

POHJOIS-KARJALAN AMMATTIKORKEAKOULU
Tietojenkäsittelyn koulutusohjelma

Hannupekka Sormunen

**ITÄ-SUOMEN AMMATTIKORKEAKOULUT -PROJEKTIN TIETO-
VARASTO**

Opinnäytetyö
Lokakuu 2012

Sisältö

1	Johdanto	6
2	Organisaatiot	6
2.1	PKAMK	6
2.2	Savonia	7
2.3	ISAT	7
3	Tietovarastot	8
3.1	Tiedon merkitys organisaatioille	8
3.2	Toiminta ilman tietovarastoa	9
3.2.1	Nykyisten järjestelmien raportit	9
3.2.2	Taulukkolaskimet	10
3.2.3	Virtuaalinen tietovarasto	11
3.3	Ongelmat ilman tietovarastoa	12
3.4	Tietovarastointi	13
3.4.1	Mitä tietovarastointi on?	13
3.4.2	Tiedon jalostus	15
3.4.3	Tietovaraston tarjoamat edut organisaatioille	15
3.5	Tietolähteet	17
3.5.1	Organisaation omat lähteet ja ulkoiset lähteet	17
3.5.2	Strukturoitu ja strukturoimaton tieto	18
3.6	Tietovarastotekniikka	19
3.6.1	Tietokannat	19
3.6.2	Tietovarastotyytit	19
3.6.3	Arkkituhtuurit	22
3.7	Tietovaraston suunnittelu	23
3.7.1	Tietovaraston laajuus	24
3.7.2	Käyttäjien tarpeet	24
3.7.3	Käytettävien tietojen valinta	25
3.7.4	Tietovaraston mallintaminen	25
3.7.5	Karkeus	26
3.8	Tietojen summaaminen ja jalostaminen	26
3.9	Tietojen reitit tietovarastoon	28
3.9.1	ETL-arkkituhtuuri	28
3.9.2	Työntömenetelmä	29
3.9.3	Vetömenetelmä	30
3.9.4	ETL:n vaiheet	31
3.10	Tietovarastöinnin työkaluohjelmistot	33
3.11	Business Intelligence	36
4	ISAT-hankkeen tietovarasto	38
4.1	Alkutilanne	38
4.2	Esitutkimus	40
4.3	Projektisuunnitelma	41
4.3.1	Johdanto	41
4.3.2	Hyödyt	42
4.3.3	Vaiheet	42
4.3.4	Organisaatio	44
4.3.5	Tavoitteet	46
4.3.6	Riskien hallinta	47
4.3.7	Aikataulu	49

4.3.8	Budjetti.....	50
4.4	Vertailu	50
4.4.1	Tietolähteet.....	51
4.4.2	Tietovarastotekniikka	52
4.4.3	Tietovaraston suunnittelu.....	53
4.4.4	Tietojen summaaminen ja jalostaminen.....	55
4.4.5	Tietojen reitit tietovarastoon.....	56
4.4.6	Yhteenveto.....	56
5	Tietovaraston hyödyt ISATille	57
6	Pohdinta.....	59
	Lähteet.....	61



POHJOIS-KARJALAN
AMMATTIKORKEAKOULU

OPINNÄYTETYÖ

Lokakuu 2012

Tietojenkäsittelyn koulutusohjelma

Länsikatu 15
80110 JOENSUU
p. (013) 260 600

Tekijä(t)
Hannu Pekka Sormunen

Nimeke
Itä-Suomen ammattikorkeakoulut -projektin tietovarasto

Toimeksiantaja
Pohjois-Karjalan ammattikorkeakoulu

Tiivistelmä

Opinnäytetyön aiheena oli tutkia Itä-Suomen ammattikorkeakoulut -projektissa toteutettavaa tietovarastoa. Projektissa oli mukana kaksi itäsuomalaista ammattikorkeakoulua, Pohjois-Karjalan ammattikorkeakoulu ja Savonia ammattikorkeakoulu. Tavoitteena oli vertailla tietovaraston toteutustapaa ja toimintamalleja yleisesti käytössä oleviin tietovaraston toteutustapoihin ja toimintamalleihin.

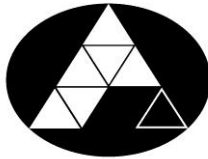
Opinnäytteessä käsitellään perustasolla tietovarastointia ja liiketoimintatiedon hallintaa. Itä-Suomen ammattikorkeakoulut -projektissa toteutettavasta tietovarastosta ja sen projektisuunnitelmasta tehdään analyysi, jossa selvitetään, kuinka tietovarastoprojekti on tarkoitus toteuttaa. Analyysissä selvitetään projektin vaiheet ja kartoitetaan tietovarastoprojektissa esiintyneitä ongelmakohtia. Näihin ongelmakohtiin tarjottiin ratkaisuja, joilla riski tietovarastoprojektin epäonnistumisesta voidaan minimoida. Analyysin perusteella laadittiin tutkimus, jossa selvitettiin tietovarasto hyödyt ammattikorkeakouluille.

Tutkimuksessa huomioitiin pääosin niitä hyötyjä, jotka tukevat ammattikorkeakoulujen liiketoimintaa. Hyödyiksi todettiin oppilaitosten tietojärjestelmien toiminnan ja kommunikoinnin tehostuminen, mikä tukee oppilaitosten operatiivista toimintaa. Lisäksi huomioitiin, että tietovaraston tarjoaman tehokkaan raportoinnin ansiosta oppilaitosten on helpompi saada selkeä kokonaiskuva toiminnastaan. Saavutettujen hyötyjen ansiosta tietovarastosta aiheutuvat kustannukset on katettu erittäin nopeasti.

Kieli
suomi

Sivuja 54
Liitteet 0
Liitesivumäärä 0

Asiasanat
tietovarastot, business intelligence, projektinhallinta

 <p>NORTH KARELIA UNIVERSITY OF APPLIED SCIENCES</p>	<p>THESIS October 2012 Degree Programme in Business Information Technology Länsikatu 15 FIN 80110 JOENSUU FINLAND Tel. 358-13-260 600</p>	
<p>Author(s) Hannupekka Sormunen</p>		
<p>Title Data warehouse of the Universities of Applied Sciences in Eastern Finland -project</p> <p>Commissioned by North Karelia University of Applied Sciences</p>		
<p>Abstract</p> <p>This thesis studies the data warehouse that was conducted by the Universities of Applied Sciences in Eastern Finland-project. The project was carried out by two universities of applied sciences that are located in the eastern Finland, the first one is North Karelia University of Applied Sciences and the second one is Savonia University of Applied Sciences.</p> <p>The purpose of this thesis was to compare the execution and approach methods of the data warehouse that was executed in the project to the commonly used data warehouse execution and approach methods. Thus data warehouses and Business Intelligence are discussed at the basic level. In addition analysis of the project plan and implemented data warehouse was made. The analysis investigates the phases of data warehouse project. Furthermore the analysis examines the problems that were encountered during the data warehouse project. Solutions to the problems that were discovered are also provided. Finally research of the benefits of the data warehouse to the universities of applied sciences was carried out on the basis of the analysis.</p> <p>It was found that data warehouse improves efficiency and communication of information systems. In addition data warehouse offers more efficient reports and analyses to both institutions. Better reports and analyses help the institutions to obtain a better understanding of their operations. In conclusion, the costs of the data warehouse are covered quickly, thanks to the benefits that data warehouse offers to institutions.</p>		
<p>Language Finnish</p>	<p>Pages 54 Appendices 0 Pages of Appendices 0</p>	
<p>Keywords data warehouses, business intelligence, project management</p>		

1 Johdanto

Tämän opinnäytetyön aiheena on tutkia Itä-Suomen Ammattikorkeakoulut - kokonaisarkkitehtuuriprojektissa (ISAT-projekti) rakennettavaa tietovarastoa. Tarkoituksena on tarkastella tietovaraston rakennetta, suunnittelua, tietovaraston rakentamista sekä projektissa käytössä olleita toimitapoja sekä menetelmiä. ISAT-projektin tietovarastoa verrataan saatujen tietojen pohjalta yleisesti käytössä oleviin tietovarastojen toteutusmenetelmiin.

ISAT-projektissa on mukana kaksi itäsuomalaista ammattikorkeakoulua, Pohjois-Karjalan ammattikorkeakoulu (PKAMK) sekä Savonia-ammattikorkeakoulu (Savonia). ISAT-projektin tietovaraston tarkoituksena on tehostaa mukana olevien ammattikorkeakoulujen tilastointia, ja mittarointia sekä helpottaa koulujen välistä tilasto- ja mittariverailua. Tietovaraston tarve syntyi koulujen tarpeesta saada kerättyä kaikki saatavilla oleva tieto yhteen, keskitettyyn paikkaan, jotta tietoa voidaan käyttää tehokkaammin ja helpommin. (Savonia 2011.) Tarvetta lisäsi ennen kaikkea uusi vuoden 2011 syksyllä voimaan astunut tietohallintolaki, jonka tarkoituksena on parantaa tietojärjestelmien yhtyeentoimivuutta ja tehostaa julkisen hallinnon toimivuutta (Valtiovarainministeriö 2011).

Opinnäytetyön toimeksiantajana toimii PKAMK. Työn tarkoituksena on perehtyä tietovarastoihin, niiden tarkoitukseen, käyttöön ja toimintaan sekä luoda kattava katsaus projektissa rakennettavaan tietovarastoon. Katsauksessa selvitetään, noudattiko tietovaraston toteutus yleisiä tietovaraston toteutusmenetelmiä.

2 Organisaatiot

2.1 PKAMK

PKAMK on Joensuun kaupungissa toimiva ammattikorkeakoulu, johon kuuluu viisi oppimiskeskusta ja kahdeksantoista koulutusohjelmaa. PKAMK:ssa opiskelee yhteensä noin 4 000 opiskelijaa ja työskentelee noin 500 työntekijää. PKAMK:n toiminta aloitet-

tiin vuonna 1992, ja sen toiminta vakinaistui neljä vuotta myöhemmin, vuonna 1996. PKAMK tarjoaa nuorille ja aikuisille korkeakoulututkintoon johtavaa koulutusta sekä osallistuu aluekehitystyön ohella myös tutkimus- ja kehittämistoimintaan. (PKAMK 2012a.)

2.2 Savonia

Savoniaa ylläpitää kuntayhtymä, joka koostuu viidestä jäsenkunnasta. Jäsenkunnat ovat Iisalmi, Kiuruvesi, Kuopio, Lapinlahti ja Varkaus. Koulutustoimintaa ei järjestetä jokaisessa jäsenkunnassa. Savonian koulutusyksiköt sijaitsevat Kuopiossa, Iisalmessa sekä Varkaudessa. Savoniassa opetetaan yli kolmessakymmenessä koulutusohjelmassa kuudella eri koulutuslallalla (Savonia 2012b). Savoniassa opiskelee noin 6 000 oppilasta ja se työllistää noin 600 työntekijää. Savonian toiminta aloitettiin vuonna 1992 ja se vakiinnutti toimintansa vuonna 1998. (Wikipedia 2012.) Savonia on yksi monipuolisimmista sekä suurimmista ammattikorkeakouluista, ja se kouluttaa vahvoja osaajia. Savonia on aktiivisesti mukana työelämässä mm. tarjoamalla palvelujaan paikallisille yrityksille. (Savonia 2012a.)

2.3 ISAT

ISAT on nimitys PKAMK:n sekä Savonian muodostamalle kumppanuudelle, jonka tehtävänä on tuottaa kilpailukykyä ja hyvinvointia suomalaisille yhdistämällä kahden ammattikorkeakoulun osaaminen ja budjetti yhdeksi suureksi kokonaisuudeksi (PKAMK 2012b). ISAT-hankkeen näkyvin puoli onkin juuri osaamisen yhdistäminen, sillä se mahdollistaa laajemman kurssitarjonnan PKAMK:n sekä Savonian opiskelijoille ristiinopiskelun myötä. Ristiinopiskelulla tarkoitetaan tilannetta, jossa esimerkiksi Savonian opiskelija suorittaa haluamansa osan opintojaan PKAMK:n tarjoamilla kursseilla. Sama käy tietenkin toisinpäin eli PKAMK:ssa opiskeleva voi halutessaan valita Savonian kursseja. Tämä käytäntö on opiskelijalle täysin maksuton ja se luo opiskelijalle mahdollisuuden oppia jotakin, mistä hän on kiinnostunut, mutta johon ei löydy opetustarjontaa oman alueen ammattikorkeakouluista. (Savonia 2012c.)

ISAT-hankkeessa tapahtuva yhdistymisen myötä koulujen tuottama tiedon määrä kasvaa, joten tiedon tehokas hyödyntäminen vaatii tehokkaampia työkaluja. Hyödyntämisellä tarkoitetaan tilastointia ja mittarointia. Halutaan esimerkiksi tietää, kuinka monta opiskelijaa kummastakin ammattikorkeakouluista valmistui yhteensä vuonna 2012. Mittaroinnilla voidaan verrata, kuinka moni valmistuneista valmistui tavoiteajassa ja onko luku erilainen muihin Suomessa toimiviin ammattikorkeakouluihin verrattuna. Tällaisten tilastoinnin ja mittaroinnin tekeminen käsin on mahdollista, mutta se vie paljon resursseja sekä on useimmissa tapauksissa todella hankalaa. Suurissa organisaatioissa sekä suurissa tietomassoissa suositellaan käytettäväksi tietokantoja ja tietovarastoja, jotka poistavat suuren määrän käsityötä sekä helpottavat tiedon hyödyntämistä monipuolisilla automaattisilla toiminnoilla. (Hovi, Koistinen & Ylinen 2001, 17.)

3 Tietovarastot

Tässä luvussa kuvataan, mitä tietovarastot ovat, kuinka ne toimivat, mihin niitä käytetään ja miten niillä voidaan tehostaa yrityksen liiketoimintaa. Tietovarastojen lisäksi tarkoituksena on luoda käsitys siitä, mitä tieto merkitsee nykyisessä liiketoiminnassa, kuinka sitä voidaan hyödyntää ja saada tehokkaammin käyttöön organisaatioissa ja yrityksissä.

3.1 Tiedon merkitys organisaatioille

Nykyisessä tietoyhteiskunnassa tiedon merkitys kasvaa jatkuvasti, ja järjestelmällisestä tiedon hallinnasta tulee yhä tärkeämpää jokaiselle organisaatiolle. Saatavilla olevaa tietoa voidaan kutsua merkitykselliseksi organisaation menestyksen kannalta vasta silloin, kun sitä voidaan hyödyntää oikein johtamisessa. (Törmänen 1999, 7.) Organisaation tallennettua sekä kerättyä tietoa tuleekin pitää elintärkeänä resurssina toiminnan kannalta, samoin kuten organisaatiossa työskentelevää henkilöstöä. Voidaankin sanoa, että organisaatiossa tehtyjen päätöksiä, ratkaisujen ja toiminnan tulee perustua kerättyyn tietoon, muussa tapauksessa yritykseltä tai organisaatiolta ei voida odottaa hyviä tuloksia. (Hovi 1997, 3–4.)

Tiedon merkityksen jatkuvan kasvun lisäksi organisaatioilla on toinenkin syy tarkkaan tiedon hallintaan, ja se on jatkuva tiedon määrän lisääntyminen. Tilannetta voidaan kuvata sanalla tietotulva. Uutta tietoa tulee jatkuvasti saataville esimerkiksi Internetin kautta. Tämä ”informaatioähky” vaikeuttaa organisaatioiden kykyä muodostaa selkeä tietokokonaisuus, sillä on vaikeaa löytää oikea, merkityksellinen tieto käytettäväksi. (Törmänen 1999, 7.)

3.2 Toiminta ilman tietovarastoa

Yritysten tiedot syntyvät operatiivisissa järjestelmissä, kuten toiminnohjausjärjestelmissä ja www-sovelluksissa. Näihin sovelluksiin syötetty tieto kertoo organisaatioiden jokapäiväisestä toiminnasta. Operatiivisiin järjestelmiin tallennettujen tietojen helposti saataville tuominen osoittautuu monelle organisaatiolle haasteeksi. Operatiivisissa järjestelmissä tiedot ovat usein hajallaan, tietorakenteet ovat hankalia ja saatavilla oleva tietoa ei ole kuvattu riittävällä tavalla. (Hovi, Hervonen & Koistinen 2009, XI.)

Tässä luvussa luodaan yleiskatsaus tilanteeseen, jossa yrityksellä ei käytössä tietovarastoratkaisua. Yleiskatsauksessa tutustutaan erilaisiin ratkaisuihin, joita organisaatiot käyttävät saadakseen yritykselle tärkeän tiedon käytettäväksi raportointiin ja analysointiin. Tarkoituksena on tuoda esiin tietovaraston hyötyjä ja selvittää, miksi tiedon hyödyntäminen ilman tietovarastoa on vaikeampaa ja resursseja kuluttavampaa.

3.2.1 Nykyisten järjestelmien raportit

Organisaatioissa on usein käytössä operatiivisia järjestelmiä, joissa on käytettävissä valmiiksi ohjelmoituja raportteja. Ongelmana valmiiksi ohjelmoiduissa raporteissa on yleisesti se, että käyttäjät sanovat haluavansa raportille kaiken tiedon. Näin he saavat erittäin laajan raportin, josta he tarvitsevat pahimmissa tapauksissa vain muutaman rivin. Valmiiksi ohjelmoiduilla raporteilla käyttäjä ei siis voi valita haluamiaan tietoja tarpeeksi tarkasti, vaan on tyydyttävä tietoihin, jotka valmiiksi ohjelmoidut raportit antavat. (Hovi ym. 2001, 22.)

Operatiivisten järjestelmien tietokannoista tehtyt raportit ja kyselyt ovat myös erittäin raskaita operaatioita, sillä niitä tehdessä tietokannoissa tehdään laajaa läpikäyntiä. Tietokannan selailusta syntyvä kuorma saattaa hidastaa operatiivisten järjestelmien toimintaa, mikä useissa tapauksissa ei ole hyväksyttävää. Kolmas ongelma on raporttien tekeminen vieraista tai hankalista tietokannoista. Raporttien tuottaminen näistä kannoista on usein hyvin hankalaa, ja työ joudutaan pahimmissa tapauksissa teettämään organisaation ulkoisella työvoimalla, mikä lisää kustannuksia. Operatiivisten järjestelmien tietokannoissa historiatiedoille ei ole omaa paikkaa, mikä huonontaa, ja pahimmillaan estää, yrityksen historia-analysoinnin tekemistä. Näiden syiden takia onkin vaikeaa ja jopa kannattamatonta yhdistää tapahtumankäsittely, raportointi sekä analysointi samaan järjestelmään. (Hovi ym. 2009, 6.)

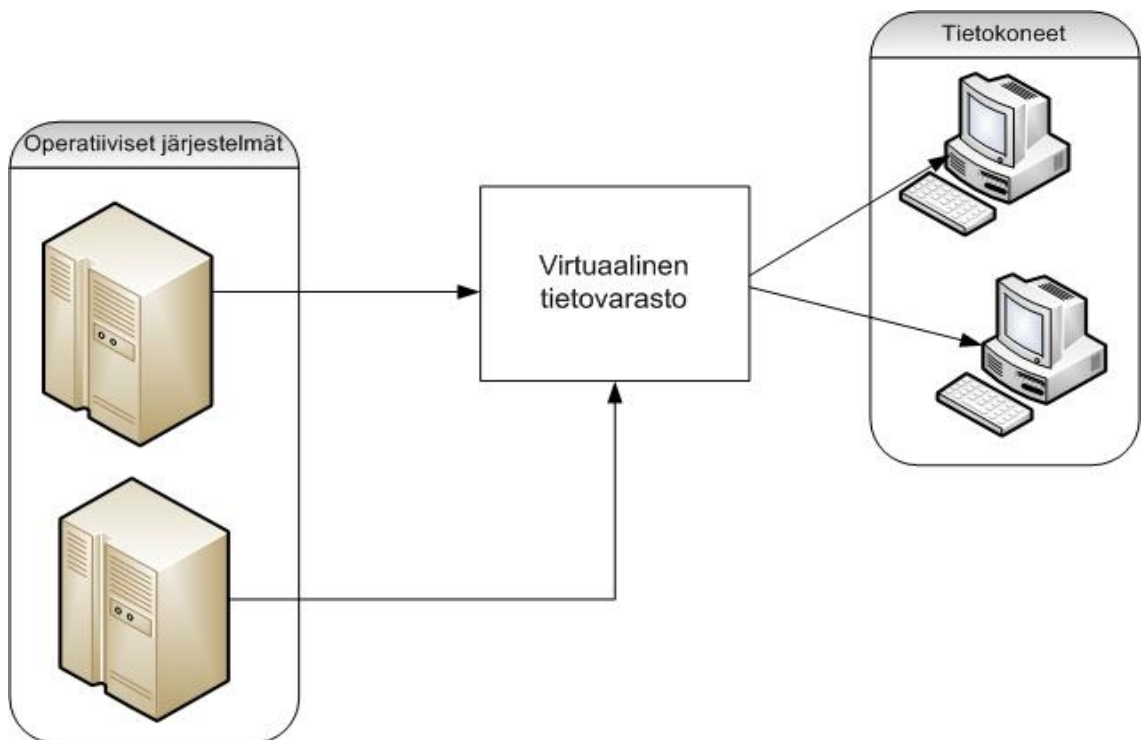
3.2.2 Taulukkolaskimet

Organisaatioissa yleinen raporttien tekoon käytettävä väline on taulukkolaskentaohjelma, kuten Microsoftin Excel tai LibreOfficen Calc. Tässä ratkaisussa suurimpana ongelmana on virhealttius, joka syntyy tietojen syöttämisestä taulukkolaskentaohjelmaan ihmisten toimesta. Tietoja myös syötetään eri järjestelmistä. Syötettyjä tietoja muokataan myös usein taulukkolaskimista löytyvillä automatisoiduilla komentosarjoilla. Tämän ”sekamelskan” seurauksena taulukkolaskentaohjelmaan rakennetusta raportointijärjestelmästä tulee hyvin monimutkainen ja vaikeasti hallittava, eikä sitä osaa ylläpitää kuin taulukon toteuttaja. (Hovi ym. 2009, 7.)

Virhealttiuden lisäksi tämä ratkaisu työllistää pahimmillaan ylimääräisiä henkilöitä tietojen syöttämiseen, vieden samalla aikaa muista tärkeistä tehtävistä, kuten tietojen analysoinnista. Taulukkolaskimia käytettäessä tietohallinnolle tärkeimmäksi tehtäväksi muodostuu erilaisten raporttien muodostus käyttäjien muuttuviin tarpeisiin. (Hovi ym. 2001, 23.) Näiden syiden takia taulukkolaskimet ovat erittäin kankeita, hitaita sekä epäluotettavia raporttien sekä analysoinnin tekoon.

3.2.3 Virtuaalinen tietovarasto

Raportteihin tarvittavat tiedot on mahdollista koota ja yhdistellä samalla kun niitä vietään raporttiin. Tämä on mahdollista virtuaalisella tietovarastolla eli tietovarastolla, jota ei ole fyysisesti olemassa, vaan tietovaraston tehtävä hoidetaan ohjelmallisesti raporttia muodostaessa (kuva 1). Tässä toimintatavassa on kuitenkin huomattavia ongelmia, sillä tietojen yhdistäminen on monimutkainen prosessi, eikä sitä voida tehdä nopeasti kuten virtuaalinen tietovarasto vaatii. Tästä syntyvää ongelmaa kuvaa tilanne, jossa pitää päätellä, kuuluvatko kahdesta tai useammasta eri järjestelmästä tulleet oppilastiedot samaan oppilaaseen, jolloin joudutaan tekemään selvitys raporttia muodostaessa. Näin raportin muodostaminen hidastuu. Kyselyn muodostaminen on hidasta jokaisella kerralla kun samantyyppinen kysely tehdään, sillä virtuaaliseen tietovarastoon ei voida tallentaa tietoa. Tällaisia tilanteita voidaan välttää selvittämällä oppilasta koskevat tiedot valmiiksi fyysiseen tietovarastoon. (Hovi ym. 2009, 8.)



Kuva 1. Virtuaalinen tietovarasto (Hovi ym. 2009, 8).

Ongelmana on myös tietojen historiointi eli säilyttäminen tulevaa vertailua varten. Sitä ei suositella tehtäväksi operatiivisissa järjestelmissä ja virtuaaliseen tietovarastoon tietoja ei voida tallentaa. Historiatiedot ovat tietovaraston tärkeimpiä ominaisuuksia ja niiden pois karsiminen vie mahdollisuuden tehdä vertailua kahden tai useamman aikajakson välillä. Ilman vertailua ei voida seurata, miten organisaatiossa tehty päätös vaikuttaa

organisaation toimintaan. Kolmantena ongelmana virtuaalisissa tietovarastoissa on operatiivisten järjestelmien ylimääräinen kuormitus, joka saattaa hidastaa järjestelmien toimintaa. (Hovi ym. 2009, 8.)

3.3 Ongelmat ilman tietovarastoa

Monissa organisaatioissa tiedon hyödyntäminen on vaikeaa ja jopa mahdotonta kankeiden operatiivisten järjestelmien takia. Nämä järjestelmät eivät tarjoa yhdenvertaista mitattavaa tietoa organisaation toiminnasta johdon käytettäväksi. (Törmänen 1999, 7.) Operatiivisten järjestelmien kehittämiseen, valmistamiseen sekä koulutukseen käytetty investointi ei tuo organisaatiolle lisäarvoa. Tästä seuraa se, että tärkeät tietoresurssit jäävät hyödyntämättä, sillä niitä ei saada käyttöön kyllin nopeasti organisaation vaihtuviin tarpeisiin. Voidaankin puhua ”tappiollisista järjestelmistä”, eli perusjärjestelmät toimivat ja ne tekevät sen, mitä niiden on alun perin suunniteltu tekevän, mutta niiden tuottamaa tietoa ei voida hyödyntää. (Hovi ym. 2001, 17–18.)

Kankeiden operatiivisten järjestelmien lisäksi ongelmana voi olla se, että hyödynnettäväksi tarvittavat tiedot voivat pahimmillaan olla täysin hajallaan eri puolilla organisaatiota. Tätä tilannetta on havainnollistettu taulukossa 1, jossa nähdään kolmen oppilaan tiedot kolmessa erilaisessa järjestelmässä. Jokaiseen operatiiviseen järjestelmään on tallennettu erilaista tietoa oppilaasta. Tilanteessa, jossa oppilaasta haluttaisiin saada kaikki organisaation operatiivisiin järjestelmiin tallennettu tieto käyttöön, jouduttaisiin tieto kokoamaan kahdesta tai useammasta järjestelmästä käsin esimerkiksi uuteen järjestelmään tai taulukkolaskentaohjelmaan. (Hovi ym. 2009, 5.) Keräystä voi hidastaa huomattavasti myös se, että tiedot ovat myös fyysisesti hajallaan. Tiedot ovat siis eri tietokoneilla ja maantieteellisesti eri paikoissa. (Hovi 1997, 7.)

Taulukko 1. Oppilaiden tiedot eri perusjärjestelmissä (Hovi ym. 2001, 20).

Student

studno	studclasskey	studname	studaddress	studphone
1333	122	Vilenius V	Honkala 2	050 123123
1338	245	Virtanen A	Honkala 5	040 123123
1472	364	Virtanen E	Honkala 2	041 123123

Oppilas

oppilasno	oppilasnimi	oppilasosoite	oppilaskunta	oppilaspuhelin	Oppilaskunnassa
1333	Vilenius V	Honkala 2	Joensuu	050 123123	Kyllä
1338	Virtanen A	Honkala 5	Kouvola	040 123123	Ei
1472	Virtanen E	Honkala 2	Kuopio	041 123123	Ei

Oppilas

oppilasno	oppilasnimi	oppilasosoite	oppilaskunta	Oppilaskunnassa
1333	Vilenius V	Honkala 2	Joensuu	Kyllä
1338	Virtanen A	Honkala 5	Kouvola	Ei
1472	Virtanen E	Honkala 2	Kuopio	Ei

Usein organisaatioissa käytettävien operatiivisten järjestelmien tietokannoissa olevaa tietoa ei ole kuvattu. Tällä tarkoitetaan tilannetta, jossa järjestelmää ylläpitävät henkilöt eivät tiedä, mitä tietokannassa olevat tietokentät tarkoittavat. Ei voida olla siis varmoja, mitä tietoja organisaation järjestelmistä löytyy ja mikä niiden tarkoitus on. (Hovi ym. 2009, 5.)

3.4 Tietovarastointi

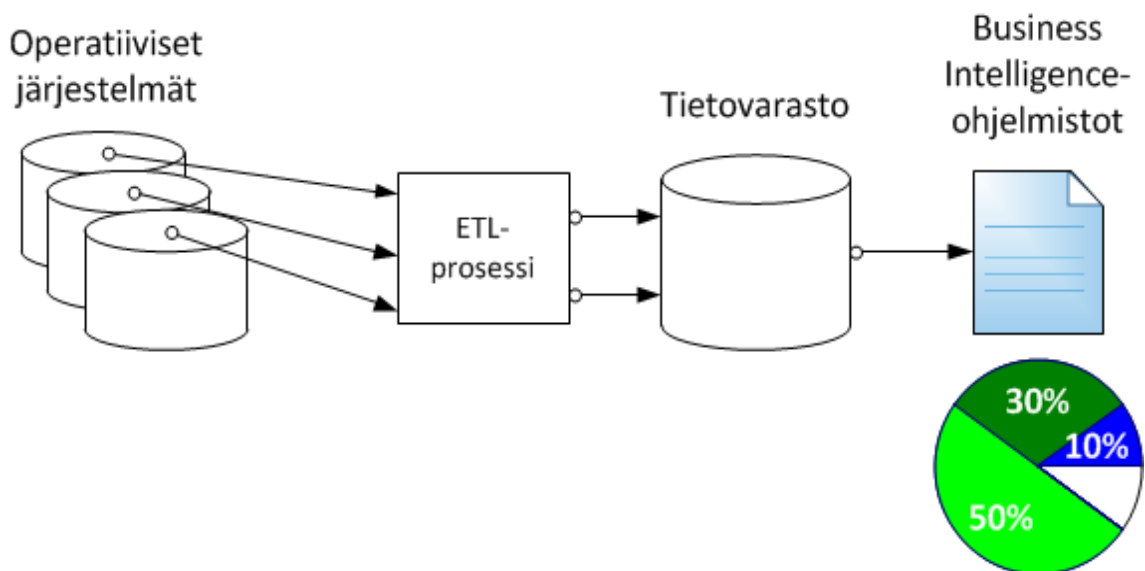
3.4.1 Mitä tietovarastointi on?

Tietovarastointi on luotu ratkaisemaan edellisessä luvussa kuvattujen organisaatioiden tiedon hyödyntämisen liittyviä ongelmia. Ratkaisuksi tiedon täydelliseen hallintaan sekä hyödyntämiseen tarvitaan keskitettyä tietokantaa, joka on suunniteltu ja toteutettu juuri tietovarastoinnin näkökulmasta. (Hovi ym. 2009, 14).

Tietovarastoon ladataan tietoa operatiivisista järjestelmistä tiettynä, ennalta määritettynä ajankohtana, yleensä kerran päivässä. Joissakin yrityksissä tiedot ladataan useimmin, joissakin harvemmin riippuen organisaation tietotarpeista. Ennen tiedon latausta se muutetaan kysely- sekä raportointikäyttöön sopivaan muotoon. Tätä muutosta kutsutaan tiedon jalostamiseksi. Tietovarastoon ladatut tiedot ovat lukukäytössä, eli niitä ei voida

muuttaa suoraan tietovarastonhallintatyökaluilla. Tämä muutosten esto varmistaa, että operatiivisten järjestelmien ja tietovaraston tiedot ovat samankaltaiset. Jos tietovaraston tietoa halutaan muokata, tulee ensin muokata operatiivisen järjestelmän tietoja. Muokkaamisen jälkeen muutettu tieto ladataan tietovarastoon seuraavan tiedon latauksen yhteydessä. Vanha tieto jää historiatiedoksi tietovarastoon. (Hovi ym. 2009, 14.)

Kuvassa 2 kuvattua prosessia kutsutaan tietovarastoinniksi. Tietovarastointi-prosessi aloitetaan operatiivisten järjestelmien tietokannoista. Näissä tietokannoissa ylläpidetään kaikki operatiivisessa järjestelmään syötetyt tiedot. Seuraavaksi siirrytään Extract, Transform, Load, -vaiheeseen eli suomeksi Pura, Muunna, Lataa -vaiheeseen eli ETL-vaihe. Tässä vaiheessa operatiivisten järjestelmien tiedot luetaan ja muokataan, eli jalostetaan tietovaraston määrittämään muotoon. Lopuksi jalostettu tieto ladataan tietovarastoon. Tietojen latauksen jälkeen tietovaraston sisältämät tiedot määritellään ja kuvataan. Kuvauksessa käytetään metatietoa. Metatiedon avulla vältetään tilanteelta, jossa tietovarastoa käyttävä organisaatio ei tiedä, mitä tietovarastoon tallennettu tieto tarkoittaa. Metatiedolla voidaan esimerkiksi kuvata seuraavia asioita tiedosta: nimi, tyyppi, käyttöoikeudet, lähdejärjestelmä ja milloin tieto on päivitetty (Hovi ym. 2009, 43). Viimeisessä vaiheessa tietovaraston sisältämää tietoa kysellään, analysoidaan ja muodostetaan raportteja Business Intelligence-työkaluilla (Hovi ym. 2009, 14).



Kuva 2. Tietojen siirtyminen tietovaraston kautta organisaation hyödynnettäväksi (Hovi ym. 2009, 43).

3.4.2 Tiedon jalostus

Tietovarastojen yhteydessä puhutaan hyvin usein tiedon selkeyttämisestä sekä jalostamisesta tietovarastoa varten. Jalostamisella tarkoitetaan tiedon yhdenmukaistamista, muuttamista samanlaiseen, ennalta määritettyyn muotoon. Jalostamista joudutaan tekemään esimerkiksi tilanteessa, jossa kahdessa tai useammassa eri operatiivisessa järjestelmässä sukupuolelle on asetettu erilaisia merkintätapoja. Järjestelmässä X sukupuoli merkitään koodeilla M ja N, kun taas järjestelmässä Y on käytössä 1 ja 0. Tietovarastoa suunnitellessa päätetään, että tietovarastossa käytetään merkintätapaa M ja N. Yhdenmukaistaminen tapahtuu siis järjestelmän Y kohdalla, jossa numero yksi muuttuu M:ksi ja numero nolla muuttuu N:ksi. Järjestelmästä X tulleita tietoja ei tarvitse jalostaa, sillä se alun perin käytti päätettyä merkintätapaa. Tällä tavalla kaikki tietovarastoon syötetyt tiedot ovat samanlaisia eikä synny tilannetta, jossa sukupuolelle onkin tietovarastosta saadusta raportissa kaksi eri merkintätapaa. Jalostus siis muuttaa tiedon helpommin luettavaksi ja ymmärrettävämmäksi. Tiedon jalostus tapahtuu kuvassa 2 esitetystä ETL-vaiheessa. (Hovi ym. 2009, 15.)

ETL-vaiheen jalostuksessa voidaan tiedon avulla tehdä myös erilaisia raportointeja tehostavia toimintoja, kuten operatiivisiin järjestelmiin syötettyjen henkilötietojen jakaminen ikäryhmiin. Jaottelu aloitetaan henkilötunnustiedon jalostuksella, eli siitä otetaan syntymäpäivä ja sijoitetaan se omaan sarakkeeseen. Lisäksi henkilön ikä lasketaan omaan sarakkeeseen. Iän muuttumisen varalta ikä lasketaan uudelleen tähän samaan sarakkeeseen jokaisella latauskerralla. Lopuksi henkilöt jaetaan ikäryhmiin ja tieto lisätään omaan sarakkeeseen. Näin raporteissa on helppoa käyttää ikäryhmiä ja kaikki tapahtuu automaattisesti tietovaraston toimesta. Ikäryhmiin jakaminen, ei ole ainoa tapa jakaa tietoa ryhmiin raportointi varten. Tietovarastoon voidaan asettaa monenlaisia automaattisia toimia raportointia helpottamaan. Tämä tekee tietovarastosta erittäin tehokkaan työkalun. (Hovi ym. 2009, 15.)

3.4.3 Tietovaraston tarjoamat edut organisaatioille

Tietovarastot ovat täysin riippumattomia organisaatiossa käytettävistä prosesseista. Tämä tarkoittaa sitä, että organisaation ei tarvitse muuttaa toimintatapojaan tietovarastoa

toteuttaessa. Tietovarasto mukautuu organisaation muuttuviin tarpeisiin. (Hovi ym. 2009, 15.)

Organisaation tuottama tieto tallennetaan sellaiseen muotoon, että siihen voidaan helposti ja nopeasti suorittaa erilaisia kyselyjä. Tieto on myös kuvattu tarkasti, joten tiedon alkuperä sekä tarkoitus tiedetään vielä useiden vuosien käytön jälkeen. Tietovarastossa voidaan säilyttää tietoa koko organisaation toiminnan ajalta. Tietovarastosta tuleekin organisaation muisti, jonne tieto on tallennettu helposti saatavilla olevaan muotoon. Organisaation tietovarastoon tallennettujen historiatietojen avulla voidaan vertailla eri aikajaksoilla tilastoituja asioita. Tätä kutsutaan aikasarja-analyysiksi. Aikasarja-analyysi on esimerkiksi oppilaitoksen oppilaiden keskiarvon vertaileminen eri vuosien välillä. (Hovi ym. 2009, 15–16.)

Riippuvuus operatiivisista järjestelmistä vähenee. Organisaation tarve käyttää samaa, jo vanhahtavaa järjestelmää poistuu, ja uuteen järjestelmään siirtyminen on paljon vaivatompaa tietovaraston ansiosta. Uusi järjestelmä liitetään ETL-prosessiin kuten vanha järjestelmä. Näin organisaatio voi jatkaa saumatta tietovarastoon tallennettujen tietojen käyttämistä uuden järjestelmän käyttöönoton yhteydessä. Näiden syiden takia tietovaraston ja operatiivisten järjestelmien yhteyttä kutsutaankin yleensä ”löyhäksi riippuvuudeksi”. (Hovi ym. 2009, 16–29.)

Tiedot ovat täysin tietovarastoa käyttävän organisaation omassa hallussa. Tiedot eivät siis jää rakennettujen järjestelmien tietokantoihin täysin käyttämättöminä. Pahimmillaan näitä tietoja päästään hyödyntämään vasta kun järjestelmän toimittaja ohjelmoi lisää raportteja organisaation käytettäväksi. Tämä maksaa yritykselle niin aikaa, kuin rahaa. Tietovarasto poistaa yrityksen riippuvuuden järjestelmätoimittajasta. (Hovi ym. 2009, 16.)

Jokainen tietoa tarvitseva löytää omatoimisesti tarvitsemansa tiedon. Tämä poistaa tarpeen luoda jatkuvasti pitkiä operatiiviseen järjestelmään ennalta ohjelmoituja raportteja, joista tarvitaan vain muutama rivi käytettäväksi. Haluttu tieto saadaan juuri oikeaan aikaan, ja juuri oikeaan tarpeeseen, käytettäväksi. Operatiivisten järjestelmien raporttien käyttöä voidaan vähentää tietovaraston raporttien ansiosta, mikä vähentää järjestelmään tulevien kyselyn määrää. Operatiivisen järjestelmien tietokannoista voidaan lisäksi pois-

taa vanhaa tietoa, joka on jo varastoitu tietovarastoon. Vanhan tiedon poisto ja operatiiviseen järjestelmään ennalta ohjelmoitujen raporttien käytön lopettaminen vähentää järjestelmän kuormitusta. Kuorman väheneminen lisää suorituskkyä eli nopeuttaa järjestelmän toimintaa. (Hovi ym. 2009, 16.)

Tietovarasto pakottaa organisaatiota parantamaan syötetyn tiedon oikeellisuutta. Tämä johtuu siitä, että tietoa voidaan tarkastella erilaisista näkökulmista, jolloin virheet paljastuvat paljon herkemmin kuin operatiivisen järjestelmien raporteista. Virheelliset sukupuoli-, ikä- tai muut tiedot, joiden perusteella oppilasta voidaan ryhmitellä, eivät välttämättä haittaa operatiivisten järjestelmien toimintaa, mutta tietovarastossa ne paljastuvat heti. Puutteellinen tieto haittaa analysointia ja tekee tietovarastosta turhan kunnes tiedon julkaisukelpoisuuteen puututaan. Tietovarasto ei siis suoraan paranna tiedon oikeellisuutta, vaan se kannustaa tiedon syöttäjiä syöttämään oikeanlaista tietoa järjestelmiin. (Hovi ym. 2009, 17.)

3.5 Tietolähteet

3.5.1 Organisaation omat lähteet ja ulkoiset lähteet

Organisaation omina tietolähteinä toimivat käytössä olevat operatiiviset järjestelmät. Näihin järjestelmiin kuuluvat esimerkiksi asiakashallintajärjestelmien ja laskutusjärjestelmien asiakas- ja taloustiedot, toimitusketjujen hallintajärjestelmien toimitustiedot ja suunnittelujärjestelmien, kuten budjetointijärjestelmistä saatavat tiedot. (Hovi ym. 2009, 18–22.)

Ulkoisia tietolähteitä ovat mm. pankki-, tilasto- sekä muut julkiset palvelut jotka tarjoavat tietoa organisaation käytettäväksi. Ulkoinen tietolähde voi olla esimerkiksi Itella tai Suomen Pankki. Nämä tietolähteet sisältävät väestö-, valuutta, ja kuntatietoja. Tietolähteiden tietoja voidaan yhdistellä organisaation omiin kyselyihin ja raportteihin. (Hovi ym. 2009, 18.) Tietoja tarjoavien palveluiden lisäksi Internetistä löytyy runsaasti tietoa organisaation käytettäväksi. Internet sisältää mm. ennusteita ja viranomaistietoja. Ulkoisia lähteitä käytettäessä tulee olla tarkkana, sillä niiden tarjoama tiedon määrä on lähes tulkoon rajaton. Ulkoisia lähteitä valittaessa tulee olla erittäin kriittinen, sillä ilman

selvää linjausta siitä, mitä tietoa käytetään, voidaan organisaatiolle aiheuttaa enemmän haittaa kuin hyötyä. Aluksi mukaan kannattaa ottaa ainoastaan sellaista tietoa, joka halutaan yhdistää omiin tietoihin. Liiallinen Internet-lähteiden käyttö voi luoda organisaatiolle tietotulvan. (Hovi ym. 2001, 77.)

3.5.2 Strukturoitu ja strukturoimaton tieto

Tietovarastoon ladataan pääosin strukturoitua tietoa. Strukturoiduksi tiedoksi kutsutaan kiinteän mittaista tietoa, joka voidaan määritellä selkeästi. Strukturoitua tietoa ovat esimerkiksi päivämäärät, nimet, opiskelijan suorittamat kurssit ja näiden kurssien arvosanat. Tiedot ovat siis mitattavissa jotenkin, esimerkiksi nimessä on tietty määrä merkkimuotoista tietoa, kun taas kurssien arvosanoissa on numeerista tietoa jne. (Hovi ym. 2009, 18.)

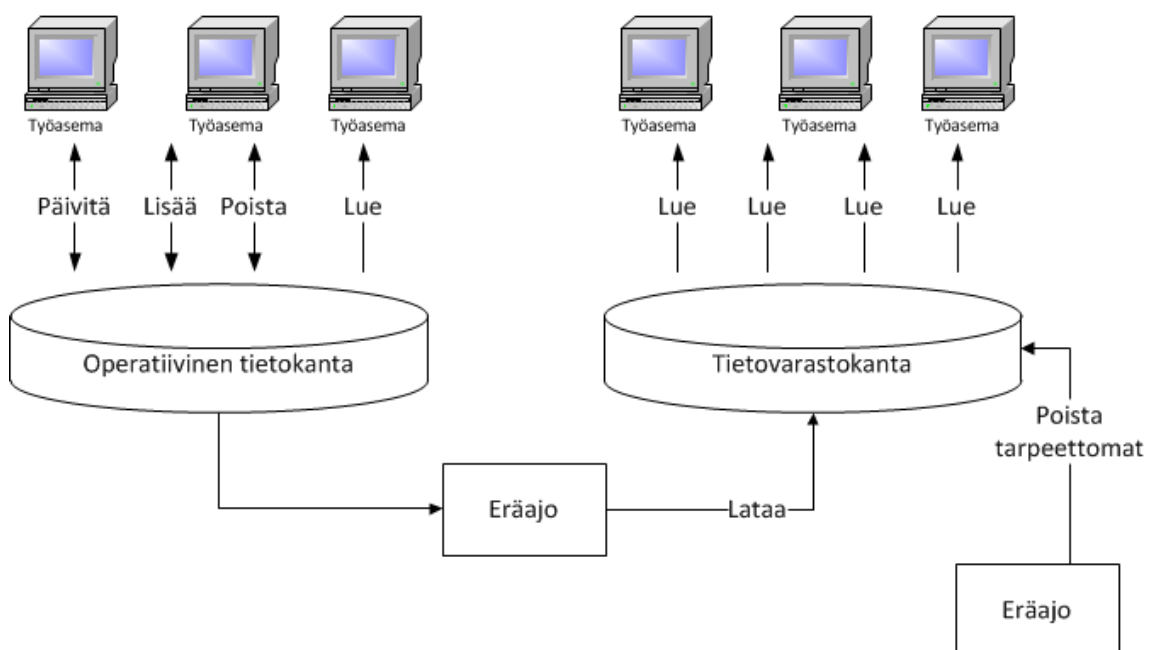
Strukturoimattomaksi tiedoksi kutsutaan dokumentteja, sähköposteja, piirustuksia, valokuvia, opetusmateriaalia yms. Strukturoimatonta tietoa ei voida mitata ja niitä useimmissa tapauksissa käsitellään omilla ohjelmillaan, esimerkiksi sähköposteja luetaan sähköpostiohjelmalla ja dokumentteja luetaan ja käsitellään tekstinkäsittelyohjelmalla. Strukturoimatonta tietoa säilytetään usein omissa tiedostoissaan tietovaraston sijaan. Mikäli tietoa halutaan siirtää tietovarastoon, joudutaan ne syöttämään tietovarastoon käsin. (Hovi ym. 2009, 18.)

Tietovarastojen kehitys on kuitenkin mahdollistamassa siirtää strukturoimatonta tietoa tietovarastoon, jolloin sitä voidaan yhdistellä muihin tietovarastoihin tallennettuihin tietoihin. Yhdistelyn ansiosta organisaatio voi esimerkiksi nähdä kaikki asiakkaalle lähetetyt sähköpostit, tekstimuotoiset sopimukset ja muut strukturoimattomat tiedot strukturoidun tiedon yhteydessä. Näin organisaatio saa laajemman näkemyksen asiakkaastaan. (Hovi ym. 2009, 18.)

3.6 Tietovarastotekniikka

3.6.1 Tietokannat

Operatiivisiin järjestelmiin syötetty tieto tallennetaan operatiiviseen tietokantaan. Näissä tietokannoissa käytetään nykypäivänä lähes poikkeuksetta SQL-pohjaisia relaatiotietokantoja. (Hovi ym. 2001, 45.) SQL-pohjaisia tietokantoja ovat mm. Oracle, SQL Server sekä DB2 (Hovi ym. 2009, 22). Tietovarastot toteutetaan useimmiten relaatiotietokantamenetelmillä eli samoilla tietokantatuotteilla kuten tietovarastot. Samanlaisesta tekniikasta huolimatta, tietovarastojen käyttö eroaa todella paljon operatiivisista tietokannoista. Kuvassa 3 esitetään erilaisia operaatioita, joita operatiiviseen tietokantoihin tehdään, kun taas tietovarastokannoista tietoa ainoastaan luetaan. (Hovi ym. 2001, 45.)



Kuva 3. Tietovarastojen ja tietokantojen eroavaisuudet (Hovi ym. 2001, 46).

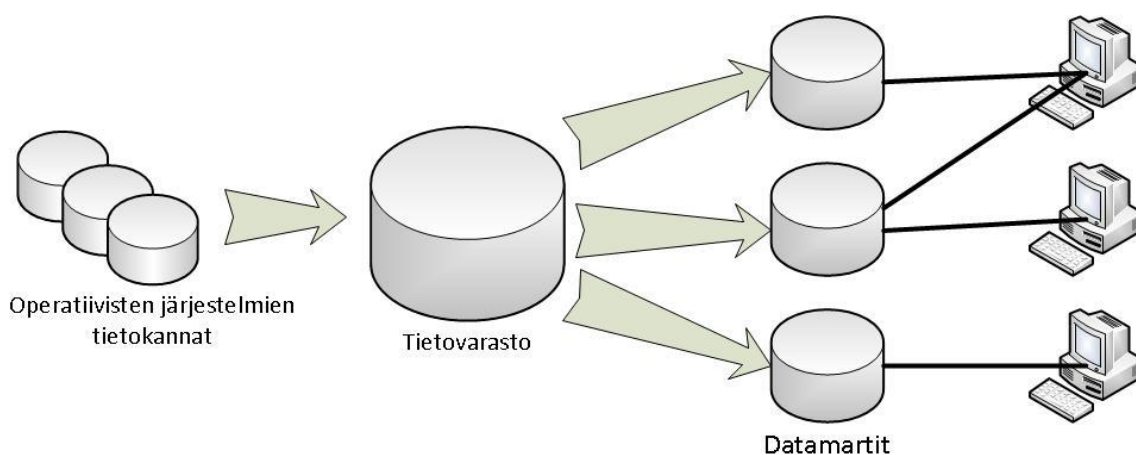
3.6.2 Tietovarastotyytit

Tietovarastoista ei ole olemassa pelkästään yhtä, kaikkeen käyttöön sopivaa tyyppiä vaan erilaisiin tarpeisiin on luotu omanlaisensa tietovarasto. Näin jokainen yritys voi valita omiin tarpeisiinsa sopivimman vaihtoehdon tai yhdistellä erilaisia tietovarastotyyppisiä niin, että se vastaa tarpeita.

Datawarehouse eli tietovarasto on useiden käyttäjäryhmien yhteiseen käyttöön tarkoitettu tietovarasto. Tietovaraston avulla voidaan tehdä kaikki tässä opinnäytetyössä kuvatut aiheet, eli varastoida ja yhdenmukaistaa tietoa organisaation käytettäväksi. Yleisesti kun puhutaan tietovarastoista, puhutaan juurikin datawarehouseista. (Hovi ym. 2009, 23–24.)

Datawarehousea pienempi tietovarasto on nimeltään datamart. Se on suunniteltu pääosin tukemaan käyttäjien raportteja ja kyselyjä. Datamartit ovat usein aihekohtaisia, kuten taloushallinnon tarpeisiin rakennettu datamart, tai organisaatiokohtaisia, kuten henkilöstöosastolle suunniteltu datamart. Datamartteja käytetään kahdella eri tavalla, ensimmäinen tapa on rakentaa vain yhden tai muutaman operatiivisen järjestelmän tietoa sisältävä datamart. Tässä tapauksessa datawarehousen ja datamartin erottaakin toisistaan tietolähteiden määrä. Siinä missä datawarehousessa voi olla satojen operatiivisten järjestelmien tiedot, sisältää datamart vain yhteen tai muutamaan järjestelmään syötetyt tiedot. Tämän lisäksi datamarttien erityisvahvuutena on niiden tuki verkon yli tapahtuvaan analysointiin sekä raporttien muodostamiseen. Datamart mahdollistaakin tilanteen jossa tietoa voidaan käyttää missä vain ja milloin vain kunhan saatavilla on verkkoyhteys. Raporttipinojen tulostelun ja mukana kantamisen voi siis unohtaa, sillä kaikki tarvittava tieto on käytössä esimerkiksi työntekijän taulutietokoneella. (Hovi ym. 2009, 24.)

Toinen tapa hyödyntää datamartteja on rakentaa johdettuja datamartteja, jolloin ne ovat isommasta datawarehousesta muodostettu, kyselyjä ja raportointia varten rakennettu tietovarastoja (kuva 4). (Hovi ym. 2009, 24.) Kuvan 4 tapauksessa datawarehousesta ei tehdä lainkaan kyselyjä, vaan kaikki kyselyt ja raportit noudetaan datamartista. Tällaisissa tapauksissa keskitetyn datawarehousen tehtäväksi jää ainoastaan yhdenmukaistaa operatiivisista järjestelmistä saapuva tieto sekä säilyttää suurempienkin organisaatioiden tuottamat massiiviset tietomäärät. (Hovi ym. 2001, 67.)



Kuva 4. Tietovarastosta muodostettuja datamartteja (Hovi ym. 2001, 67).

Tilannekannaksi (Operational Data Store) kutsutaan ajantasaista tai lähes ajantasaista tietokantaa, joka yhdistelee usean operatiivisen järjestelmän tietoja. Tilannekannassa tietoa ei kuitenkaan jalosteta tai historioida kuten datawarehouseessa. Tilannekanta eroaa muista tietovarastotyypeistä siten, että sen sisältämää tietoa voidaan muokata, eli tieto ei ole ainoastaan luku käytössä. Tilannekannan etuna onkin se, että jo yhdisteltyä tietoa voidaan muokata ennen kuin se ladataan datawarehouseen. (Hovi ym. 2009, 25.) Tilanteita, joissa tilannekantaa tarvitaan, ovat mm. vanhoista operatiivista järjestelmistä tulneiden tietojen muokkaus muilla sovelluksilla. Vanhojen operatiivisten järjestelmien tietokannat ovat usein erittäin hankalia rakenteeltaan. Pahimmillaan järjestelmän toimittajalle joudutaan tekemään toimeksianto jokaisella kerralla kun tietoa halutaan siirtää tietovarastoon. Tällöin tietojen siirto ja muokkaus on erittäin vaikeaa ja kallista. Tällaisissa tapauksissa tietojen lataus suoraan tilannekantaan ja tiedon muokkaaminen toisella sovelluksella on yksinkertaisempi ja halvempi vaihtoehto. (Hovi ym. 2001, 65.)

Työalueeksi kutsutaan tietokantaa tai joukkoa tietokantoja joita käytetään tiedon latauksen työtilana. Usein tietojen latausvaiheet toteutetaan niin, että operatiivisten tietojärjestelmien tiedot luetaan työalueelle odottamaan käsittelyä ja tietovarastoon vientiä. Työalue on hieman kuin tietovaraston lastauslaituri, sille tuodaan tieto joka myöhemmin kuljetetaan ETL-prosessia hyväksikäyttäen tietovarastoon. Käyttäjät eivät pääse käyttämään työalueen tietoa. (Hovi ym. 2009, 25.)

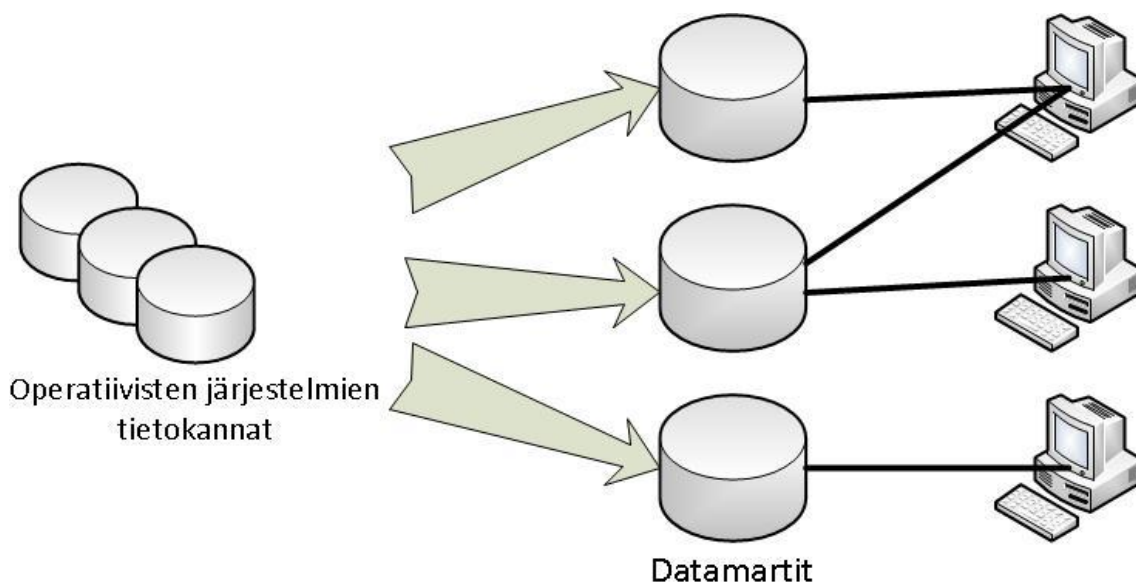
3.6.3 Arkkitehtuurit

Edellisessä luvussa esiteltiin erilaisia tietovarastotyyppejä ja samalla todettiin, että jokaiselle tietovarastoa käyttävälle organisaatiolle on omanlaisensa ratkaisut. Tietovarastotyyppien lisäksi tietovarastoja voidaan rakentaa erilaisilla arkkitehtuureilla. Käytännössä erilaisia arkkitehtuurivaihtoehtoja on kolme. Nämä vaihtoehdot ovat yksi tai useampi erillinen datamart, keskitetty yritystason tietovarasto sekä joukko yhdenmukaistettuja datamartteja. (Hovi ym. 2009, 25–26.)

Erilliset datamartit -ratkaisussa rakennetaan yhden tai muutaman sovelluksen tiedon tallentamista varten. Tällainen arkkitehtuurin on suunniteltu pieniin organisaatioihin tai yrityksiin, joissa tietovaraston käyttäjämäärät ovat pieniä ja aihealueet suppeita. Tällaisia ovat esimerkiksi talous- tai henkilöstöhallinto. Etuna tässä arkkitehtuurissa on sen erittäin nopea toteutus, jolloin raportit saadaan käyttäjille hyvinkin nopeassa ajassa. Nopeuden lisäksi hyvillä Business Intelligent -työkaluilla datamartista saadaan jaettua tietoa tehokkaasti käyttäjille esimerkiksi Online-käyttöön Internetin avulla. Ongelmana tässä ratkaisussa on juuri tietojen erillisuus. Yritys ei saa rakennettua kokonaiskuvaa asiakkaasta, sillä jokaisesta asiakkaasta ladataan vain tarvittava osa datamarttiin. Tämä tarkoittaa sitä, että esimerkiksi tilaustietoja sisältävään datamarttiin tulee ainoastaan tilaustietoja. Samoja tietoja saatetaan ladata useita kertoja eri datamartteihin, joka lisää operatiivisiin järjestelmiin muodostuvaa raskautta. Näiden syiden takia suurempien organisaatioiden tai yritysten ei kannata tuhlaa resursseja useiden erillisten datamarttien rakentamiseen. (Hovi ym. 2009, 26.)

Keskitetyssä yritystason tietovarastossa, eli Enterprise Data Warehousessa (EDW) idea on koota ja yhdistellä kaikki organisaation tiedot yhteen tai muutamaan isoon tietokantaan. Tietokanta pitää sisällään usean liiketoiminnallisen alueen tiedot yhdenmukaistettuna jolloin tietoa voidaan tarkastella yritystasolla, yli sovellus- ja organisaatorajojen. Yhdestä asiakkaasta saadaankin EDW:n ansiosta paljon parempi kokonaiskuva kuin erillisillä datamarteilla tai tilanteessa, jossa tietovarastoa ei ole käytetty ollenkaan. EDW:stä voidaan tehdä raportteja sekä muodostaa summatauluja ja johdettuja datamartteja. Summataulujen ja johdettujen datamarttien rakenteet suunnitellaan mahdollisimman helppoiksi. Helppouden ansiosta tietovarastoa käyttävien työntekijöiden on yksinkertaista tehdä kyselyjä ja analysoida tietovarastoon tallennettua tietoa. (Hovi ym. 2009, 26.)

Yhdenmukaistetut datamartit-arkkitehtuurissa rakennetaan yhden suuren tietovaraston sijaan useita datamartteja, jotka sisältävät osittain yhtenäistä tietoa toistensa kanssa (kuva 5) (Hovi ym. 2009, 28). Tämä ratkaisu on ns. bottom-up -lähestymistapa, jossa ensin rakennetaan yksi datamartti ja siitä laajennetaan tarpeen mukaisesti uusiin datamartteihin. Organisaatiolla voi olla esimerkiksi käytössään ainoastaan myyntitietoja sisältävä datamart ja myöhemmin huomataan, että tarvitaan asiakastietoa sisältävä datamart. Tässä ratkaisussa tulee olla erittäin tarkkaavainen, että jokaisessa datamarttia käytetään samasta asiasta samaa nimitystä, eli rekka on rekka jokaisessa datamartissa. Muuten joudutaan samanlaiseen ongelmaan kuin usean operatiivisen järjestelmän kanssa, eli tieto on hajallaan ja sitä ei voida yhdistellä. (Nishith 2006.)



Kuva 5. Yhdenmukaiset datamartit (Hovi ym. 2001, 68-69).

3.7 Tietovaraston suunnittelu

Tietovarastoprojektin onnistumisen kannalta tarkka suunnittelu on erittäin tärkeää, sillä suunnittelu on yksi tietovaraston laajempia osa-alueita. Suunnittelun puutetta tai kehnoutta eivät pelasta hyvät työkalut tai tietovarastotuotteet. Tietovaraston suunnittelussa on kolme tärkeää kohtaa, jotka tulee ottaa huomioon: laajuus, tarpeet ja saatavilla oleva tieto. (Hovi ym. 2009, 31.)

3.7.1 Tietovaraston laajuus

Tietovarastoa toteuttaessa tulee aina ensimmäisenä ottaa huomioon kohdealueen laajuus. Mikäli mallinnetaan ainoastaan oppilashallintoa, kuuluu mukaan ottaa mm. oppilaasta, kursseista ja opintopisteistä kertovaa tietoa. Tällaisessa tilanteessa on turhaa alkaa rakentamaan kokonaisen yrityksen kattavaa EDW:tä, josta saadaan tietoa myös henkilöstöhallinnon sekä ostotoiminnan tietoja, vaan tehdä yksittäinen datamart joka keskittyy rajattuun alueeseen. Datamarteissa laajuuden huomioon ottaminen on erittäin tärkeää, sillä niitä ei ole suunniteltu suuri tietomassoille kuten EDW:t. Tästä huolimatta EDW-ratkaisussakin tulee huomioida laajuus, sillä päämäärätön tietojenlataus tietovarastoon ei tuo yritykselle lisäarvoa. Ennen EDW:n luontia onkin hyvä luoda visio siitä, mitä siltä odottaa ja miten käyttäjät voivat hyödyntää omassa organisaatiossaan. (Hovi ym. 2009, 32.) Oman organisaation tietojen lisäksi tulee päättää, minkälaisia ulkoisia tietoja tietovarastoon otetaan mukaan. Päätöksiä tulee esimerkiksi tehdä siitä, kuinka monen kilpailijan tiedot tietovarastoon ladataan sekä mitkä tilastotiedot ovat välttämättömiä. Ilman selkeää rajausta tietovarastoon saatetaan ladata suuria määriä hukkatietoa, joka on täysin käyttämättömän viemässä tilaa tarpeelliselta tiedolta ja aiheuttamassa tietotulvaa. (Hovi 1997, 67.)

3.7.2 Käyttäjien tarpeet

Usein tietoa hyödyntävillä henkilöillä saattaa olla todella selkeitä näkemyksiä siitä, mitä tietoja halutaan onnistuneiden raporttien ja analyysien tekemiseen. Tietovarastoa suunnitellessa kannattaakin ottaa huomioon käyttäjien näkemyksiä. Joissakin tapauksissa käyttäjät eivät tunne tietotarpeita ja saattavat vaatia yksinkertaisesti kaiken mahdollisen tiedon mitä suinkin voidaan kerätä. Tällaisissa tapauksissa kannattaa tehdä analyysi käyttäjien tarpeista esimerkiksi haastatteluilla. Tarpeiden selvittämisestä huolimatta on erittäin vaikea ennustaa, mitä tietoa käyttäjät kahden vuoden kuluttua tarvitsevat. Kehityssuuntana nykyisissä EDW-ratkaisuissa on, että operatiivisista järjestelmistä ladataan varmuuden vuoksi suurin osa tiedoista. (Hovi ym. 2009, 32.)

3.7.3 Käytettävien tietojen valinta

Kohdealueen määrittelyn ja käyttäjien tarpeiden selvittämisen jälkeen tulee päättää, mitkä tiedot operatiivisista järjestelmistä otetaan mukaan ja mitkä tiedot jätetään tarpeettomana lataamatta tietovarastoon. Tietojen poiminta on välillä erittäin vaikeaa, sillä koskaan ei voida tietää, mitä tietoa käyttäjät saattavat tarvita tulevaisuudessa. Tähän ongelmaa on muutama ratkaisu. Ensimmäinen tapa on ladata kaikki tieto mitä organisaatiossa tuotetaan. Nykypäivänä tämä on hieman halvempaa kuin esimerkiksi kymmenen vuotta, sillä levytila on koko ajan halpenemassa. Toinen tapa on poimia mukaan kaikki nykyisiä tietotarpeita vastaavat tiedot ja toteuttaa tietovarastoprojekti mahdollisimman nopeasti. Tietotarpeiden muuttuessa tehdään laajennuksia poimintoihin ja toteutetaan projekti jälleen mahdollisimman nopeasti. Näin saadaan lähestulkoon aina kaikkia tietotarpeita tyydyttävä poiminta ja samalla todellisia kokemuksia siitä, mitä tietoja todella tarvitaan. (Hovi ym. 2009, 33.)

3.7.4 Tietovaraston mallintaminen

Täysin suunnitelmien mukaisesti toimivaan tietovarastoon johtava projekti on pitkä ja se saattaa pisimmällään kestää useita vuosia. Tästä huolimatta ei ole hyväksyttävää, että tietovarastoa päästään hyödyntämään vasta sen valmistuttua. Tietovarastoprojektin tavoitteena on tehdä mahdollisimman laaja, operatiivisten järjestelmien tietoja yhdistelevä tietovarasto nopeasti käytettäväksi. Tämä mahdottomalta kuulostava yhtälö voidaan tehdä mahdolliseksi mallintamisen avulla. Mallin luominen aloitetaan luomalla ylätasoin kokonaismalli, joka kattaa koko tietovaraston. Tällä lailla saadaan hahmotettua kokonaiskuva tietovarastosta.

Seuraavaksi suunnitellaan ensimmäinen toteutettava osa-alue, joka perustuu käyttäjien asettamiin tietotarpeisiin ja operatiivisten järjestelmien tietoihin. Nopeiden tuloksien takaamiseksi suunnitelman valmistuttua aletaan välittömästi rakentaa ensimmäisen vaiheen tietovarastoa käyttäjille. Lopuksi suunnitellaan ja toteutetaan seuraava suunniteltu osa-alue. Tätä viimeistä vaihetta toistetaan niin kauan, kunnes jokainen osa-alue on toteutettu. Tällaista mallia hyödyntämällä taataan nopea tyydytys tietotarpeisiin ja samalla saadaan kehitettyä suurempaa kokonaisuutta. Nopea toteutus myös lisää tietovarastoin-

tiprojektiin uskottavuutta joka on erittäin tärkeää pitkissä hankkeissa. (Hovi ym. 2009, 33–34.)

3.7.5 Karkeus

Karkeisuudella tarkoitetaan tietovarastoon vietävän tiedon yksityiskohtaisuutta. Hienojakoisimmilla karkeisuuden tasolla tiedot viedään tietovaraston sellaisena, kuten ne on syötetty operatiiviseen järjestelmään. Tämä tarkoittaa sitä, että esimerkiksi elintarvikekaupan tietovarastoon viedään jokainen kassatapahtumarivi. Tästä tiedosta voidaan erottaa mitä kaupasta ostetaan maanantaisin kello 10–12. Samalla voidaan tutkia mitä asiakkaat ostavat, eli tutkia asiakkaiden ostoskäyttäytymistä tiettyinä ajanjaksoina. Ainoa ongelma jokaisen kassatapahtumarivin tallentamisessa on tietueiden suuri määrä, joka taas tarkoittaa suurempaa tietokantaa ja suurempia levykapasiteettikustannuksia. Nykyinen trendi on tallentaa tiedot erittäin tarkalla tasolla, jotta saadaan kaikki mahdollinen irti tietovaraston tiedoista. (Hovi ym. 2009, 34.)

Vastakohta hienojakoisuudelle on erittäin karkea tiedon yksityiskohtaisuus. Tällaisissa tilanteissa tallennetaan ainoastaan päiväkohtaista myyntiä, eli esimerkiksi kuinka paljon maitoa on myyty päivässä. Tällöin ei saada niin tarkkoja tietoja kuin hienojakoisella tasolla, mutta levykapasiteettia käytetään vain murto-osa. Vaihtoehdot eivät silti ole pelkästään näiden kahden välillä vaan näitä kahta menetelmää voidaan myös yhdistellä, eli seurataan myynti päivätasolla, mutta tallennetaan kuittirivitaso esimerkiksi kuukauden ajalta, eli ns. otos. Tällä tavalla päästään tekemään tarkempia analyyskejä, mutta tietovaraston tietokanta voidaan pitää pienempänä. (Hovi ym. 2009, 34.)

3.8 Tietojen summaaminen ja jalostaminen

Tiedon summaus ja jalostaminen ovat tyypillisiä ja erittäin olennaisia menetelmiä tietovarastoissa. Summausta ja jalostamista käytetään kyselyjen helpottamiseksi ja hyvän suorituskyvyn takaamiseksi. Ennen kuin tietoa voidaan alkaa summata ja jalostaa tulee se tallentaa tietovaraston tauluihin. Näitä tarkalla tasolla sijaitsevia tauluja kutsutaan perustauluiksi. Perustauluihin tallennetaan, esimerkiksi kaikki myyntitapahtumat, joten perustauluista tulee usein erittäin isoja. Suurten tietomassojen takia perustauluista kyse-

leminen on hyvin hidasta. Perustauluista muodostetaankin summatauluja juuri kyselyä nopeuttamaan, sillä ne sisältävät paljon vähemmän rivimääriä kuin perustaulut. Summatauluissa on usein mukana myös sellaista tietoa, jota voidaan helposti seurata, kuten esimerkiksi päivämääriä (dimensiotietoa). Summataulutasoille suositellaan myös laskettavaksi valmiiksi johdettuja tietoja, eli erilaisilla laskukaavoilla perustiedoista jalostettua tietoa. Näitä tietoja ovat mm. asiakaskate ja kannattavuus. Summataulut ovat arkkitehtuurisesti aivan samanlaisia kuin datamartit. (Hovi ym. 2009, 44.)

Usein samoja tunnuslukuja ja summia tarvitaan useihin raportteihin ja kyselyihin, eikä niitä kannata laskea aina uudelleen jokaiselle raportille. Summatauluihin voidaankin laskea toistuvat tiedot valmiiksi, jota kaikki käyttäjät voivat käyttää. Summataulujen ansiosta yrityksessä toimivien organisaatioiden on helpompi saavuttaa yhteinen totuus, eli tilanne, jossa tunnusluvut lasketaan yhteisesti hyväksytyllä kaavalla ja jota jokainen tietovarastoa käyttävä henkilö käyttää. Ilman tietovarastoa voidaankin joutua pahimmillaan tilanteeseen, jossa kahdessa eri raportissa on samasta tunnusluvusta erilainen tulos, esimerkiksi yrityksen katteesta. (Hovi ym. 2009, 45.)

Summatauluista on lisäksi vielä yksi variaatio, jota kutsutaan yhteenvetosummatauluksi. Muihin tauluihin erona on se, että yhteenvetosummataulu on datamart, jonka tarkoituksena on viedä mahdollisimman vähän levytilaa. Yhteenvetosummataulun tarkoituksena on täyttää käyttäjien akuutit tietotarpeet. Yhteenvetotauluun viedään mahdollisimman paljon summattietoa, joka kiinnostaa käyttäjiä, ja jolle on käyttöä lähes päivittäin. Lisäksi yhteenvetotauluun otetaan mukaan johdettua tietoa ja historiasummattietoja. Pienen levytilan takia historiatietoa ei kuitenkaan ole paljoa. Lisäksi historiatieto on jo pitkälle summattu, kuten esimerkiksi vuoden 2006 liikevaihdon ero vuoden 2007 liikevaihtoon. Ehdoton etu yhteenvetosummatauluissa on niiden nopeus, joka tulee juuri pienestä koosta. Haittana on tietysti tiedon suppeus, eli jos halutaan tarkempaa tietoa, esimerkiksi analysointia varten, käytetään keskitetyn tietovaraston tietoja, joka ei ole valmiiksi summattuna. (Hovi ym. 2009, 46–47.)

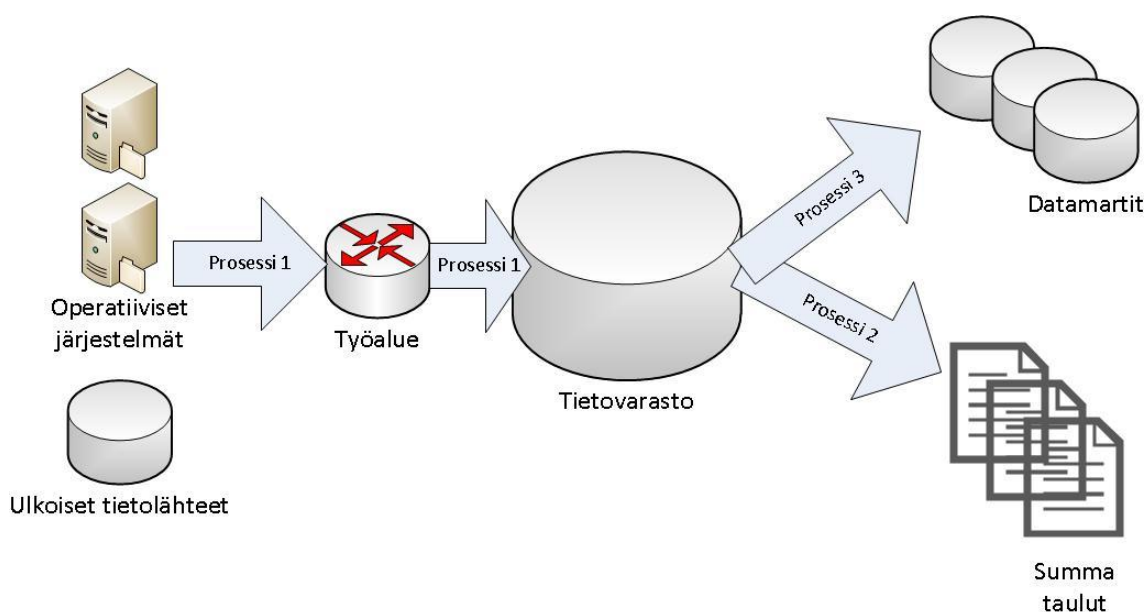
Summaamisella voidaan ratkaista mm. seuraavanlainen suorituskyvyllinen ongelma. Yrityksellä on myynnin seurannan tietovarastossa yli 20 miljoonaa riviä tuotetietoja. Tuoteryhmäkohtainen analyysi tästä tietovarastosta kestää todella kauan, joten suorituskyvyn lisäämiseksi tietovarastoon luodaan summataulu. Tämä summataulu sisältää tuo-

teanalyysin tarvittavat tiedot ilman asiakastietoja, sillä tuoteanalyysiin ei tarvita asiakastietoja. Ilman asiakastietoja summatauluun tulee ainoastaan n. 120 000 riviä. Nyt tuoteanalyysiin tarvittava kyselyt voidaan tehdä tästä summataulusta, jolloin analyysin luominen ei kestä niin kauan ja tietovarastoon syntyvä kuorma on paljon pienempää. (Hovi ym. 2001, 104.)

3.9 Tietojen reitit tietovarastoon

3.9.1 ETL-arkkitehtuuri

Aikaisemmassa luvussa sivuutettiin ETL-vaihetta, jota käytetään tiedon poimintaan, muokkaamiseen ja lataukseen tietovarastoon. Tämä ei kuitenkaan ole ainoa käyttökohde ETL-arkkitehtuurille, vaan sitä voidaan käyttää kaikkeen tiedon siirtämiseen, kuva 6. Kuvassa 6 esitetään erilaisia ETL-prosesseja. Prosessissa yksi kuvataan raakatiedon poimintaa operatiivisista järjestelmistä ja ulkoisista lähteistä, tietojen muokkaamista ja lataamista tietovarastoon. Prosessissa kaksi kuvataan summataulujen muodostusta tietovaraston perustauluista. Prosessissa kolme muodostetaan datamartit analysointia ja raportteja varten. Näiden kolmen prosessin saumaton toiminta on erittäin tärkeää, sillä kun tietojen ajastettu latausprosessi alkaa suoritetaan prosessit mainitussa järjestyksessä. ETL-prosessissa on käytössä erillinen työalue, jota käytetään apuna tiedon muokkauksessa ja jalostuksessa. (Hovi ym. 2009, 48.)



Kuva 6. ETL-prosessit (Kerr 2003).

ETL-prosessien toteuttaminen on tietokantapohjaisten eräajosovelluksien ohjelmointia. Tähän ohjelmointiin käytetään usein ETL-työkaluja, jotka on suunniteltu helpottamaan ja tehostamaan työtä. ETL-työkalut ovat eräänlaisia sovelluskehittäjiä, joissa on valmiiksi rakennettuja osia tyypillisiin lataustilanteisiin. Lisäksi tuotteissa on usein mukana graafinen suunnittelutyökalu, jolla voidaan suunnitella tietojen siirtymiset ja muunnokset. Työkalu lisäksi muuttaa kaiken suunnitellun käyttövalmiiksi koodiksi. Työkalujen tuomasta helpotuksesta huolimatta ETL-prosessin suunnittelu, testaus ja toteuttaminen vievät 60–80% tietovarastonhankkeen toteuttamiseen käytettävästä ajasta. ETL-prosessin toteuttaminen oikein onkin tietovarastohankkeen onnistumisen kannalta erittäin tärkeää. (Hovi ym. 2009, 48–54.)

3.9.2 Työntömenetelmä

Tiedon lataukseen on käytössä kaksi perusvaihtoehtoa, ensimmäisenä käsittelyssä on työntömenetelmä. Työntömenetelmässä operatiivisissa järjestelmissä sijaitsevat tiedot poimitaan, muokataan sovittuun muotoon ja kirjoitetaan tiedostoihin. Työntömenetelmän nimi tulee siitä, että operatiivinen järjestelmä ”työntää” tiedot ulos tiedostoihin. Tiedostot usein siirretään tietovarastokoneelle eri ympäristöön. Siirron jälkeen tiedostot luetaan sisään, muokataan vielä kerran tietovarastoa varten. Muokkaamisen yhteydessä samanlaista tietoa yhdistellään, ettei tietovarastoon kirjoiteta jokaisesta järjestelmästä samaa tietoa. Lopuksi tiedot kirjoitetaan tietovarastoon. Ensimmäisen vaiheen poiminnassa vastuu on operatiivisen sovelluksen ylläpitäjällä. Useimmiten tämä ylläpitäjä on organisaation ulkopuolinen henkilö, etenkin silloin kun kyseessä on järjestelmä, joka on ostettu ulkoiselta toimittajalta. Toisen vaiheen siirron rakentavat ETL-toteuttajat. Tällä tavalla operatiivisen järjestelmän ja tietovaraston väliin syntyy selkeä rajapinta. (Hovi ym. 2009, 50–51.)

Etuna tässä menetelmässä on joustavuus. Organisaation ei tarvitse käyttää samaa operatiivista järjestelmää, vaan se voidaan vaihtaa tarpeen vaatiessa. Uudelta toimittajalta tilataan samanlaiset siirtotiedostot kuin korvattavassa järjestelmässä ja tietovaraston käyttö voi jatkua ilman häiriöitä. Järjestelmän vaihtamisen helppous on todella tärkeää useimmille organisaatioille, sillä joissakin tilanteissa hyödyttömästä järjestelmästä on tärkeää päästä helposti eroon. Lisäksi ulkoisten toimittajien asiantuntevuus ensimmäisen vaiheen poiminnassa on suuri etu, sillä operatiivisten järjestelmien tietokannat voivat

olla hyvin vaikeaselkoisia. Joissakin tilanteissa työntömenetelmä on ainoa tapa saada tieto siirrettyä järjestelmistä tietovarastoon, sillä jotkut järjestelmätoimittajat eivät anna operatiivisen järjestelmänsä tietovaraston kuvauksia ja tietomalleja yrityksien käytettäväksi. Ilman kyseisiä tietoja, tietojen lataus suoraan tietokannasta on lähestulkoon mahdotonta. Vaikka yrityksellä olisi jonkinlainen käsitys järjestelmän tietokannasta ja sen rakenteesta, tietokannan toimintaa ei välttämättä ole kovin helppo ymmärtää. Tällöin on parasta tilata toimittajalta yhdessä sovitut tiedostot, jolloin vastuu syntyvistä tiedostoista ja niiden laadusta on toimittajalla. (Hovi ym. 2009, 51.)

Työntömenetelmässä tietojen lukemisen aloite tulee operatiivisista järjestelmistä, jolloin poimintaan sopiva hetki voidaan ajoittaa tarkasti. Tällöin ETL -prosessi ei lue väärään aikaan tietoja, kuten esimerkiksi päivityseräajon aikana. Menetelmässä käytettyjen tiedostojen luoma rajapinta tarjoaa etuja lataukseen. Rajapinnan ansiosta esimerkiksi katkenneen latausajon uudelleen suorittaminen on paljon helpompaa kuin suoraan tietokannoista lukiessa. Katkennut lataus voidaan aloittaa tiedoista heti ilman, että tietoihin on tullut muutoksia. Tietokannoista suoraan lukiessa voi tilanne olla jo menetetty muutuneiden tietojen takia. Tiedosto luovatkin ns. ”turvapaikan” ladattavalle tiedolle. Näitä tiedostoa voidaan lisäksi tutkia erilaisten virhetilanteiden sattuessa, jolloin saatetaan saada tieto siitä, miksi virhetilanne tapahtuu. (Hovi ym. 2009, 51.)

Työntömenetelmässä on myös ongelmia, sillä tiedostojen lukeminen ETL-työkaluilla on usein paljon hankalampaa kuin suoraan tietokannasta lukeminen. Tämä johtuu siitä, että tietokannan tietokuvauksien, kuten nimien ja tietotyyppien, siirtäminen tiedostoon ei ole niin helppoa kuin suoraan tietokannasta siirrettäessä. Lisäksi työntömenetelmässä on paljon enemmän vaiheita sekä mahdollisia kustannuksia kuin suoraan tietokannasta luettaessa. (Hovi ym. 2009, 51.)

3.9.3 Vetomenetelmä

Vetomenetelmässä ei käytetä ulkoisia tiedostoja vaan tiedot poimitaan ja ladataan suoraan operatiivisen järjestelmän tietokantojen tauluista. Tämä onnistuu samalla ETL-arkkitehtuurilla, jota käytetään työntömenetelmässä. Tiedostokytkösten sijaan ETL-työkalut ohjelmoidaan kytkeytymään suoraan tietokantaan, jolloin tietojen nimet ja kuvaukset saadaan siirrettyä suoraan ETL-työkaluun. Etuna tässä menetelmässä on sen

yksinkertainen toteutus, joustavuus sekä nopeus. Nämä edut tosin tavoitetaan vain silloin, kun operatiivisissa järjestelmissä käytettävät tietokannat tunnetaan erittäin hyvin. Mikäli tietokantaa ei tunneta, on vaarana, että sattuu kohtalokas väärinymmärrys, joka aiheuttaa virhetilanteen niin tietovarastolle kuin tietokannalle. (Hovi ym. 2009, 52.)

Ongelmia vetomenetelmässä on paljon enemmän kuin työntömenetelmässä. Vetomenetelmää käytettäessä on vaarana, että tietovarastoon siirretään keskeneräisiä tietoja, esimerkiksi kesken järjestelmäpäivitysten. Tällaisen virheen sattuessa on uusintalatauksen suunnittelu ja toteutus paljon monimutkaisempaa kuin työntömenetelmässä. Pahimmillaan joitakin tietoja ja versioita tiedoista ei saada enää koskaan ladattua tietovarastoon, sillä ne ovat jo muuttuneet tai poistuneet tietokannasta. Vetomenetelmä ei ole samalaista rajapintaa tietolähteiden ja tietovaraston välillä, kuten työntömenetelmässä. (Hovi ym. 2009, 52.)

ETL-toteuttajille syntyy eniten työtä vetomenetelmää käyttäessä, sillä operatiivisen järjestelmän vaihtuessa he joutuvat toteuttamaan koko ETL-prosessin uudelleen. Lisäksi ETL-toteuttajat eivät välttämättä ymmärrä täysin operatiivisen järjestelmän tietokannan tietorakenteita ja merkitystä, jolloin oikeaa tietoa ei saada ladattua tietovarastoon. Tämä aiheuttaa lisätyötä sekä kuluja aina järjestelmää vaihtaessa. (Hovi ym. 2009, 52.)

3.9.4 ETL:n vaiheet

Ensimmäinen vaihe ETL:ssä on tietojen poiminta. Tietojen poiminnassa käytetään yleisesti samoja toimintaperiaatteita, joita käsiteltiin aikaisemmin työntö- ja vetomenetelmien yhteydessä. Poiminta on siis parhaimmillaan suoraviivaista tietojen lukua operatiivisen järjestelmän tietokannan tauluista. Joissakin tapauksissa tiedon jalostamista voidaan tehdä jo tietojen poiminnan yhteydessä. (Hovi ym. 2009, 55.) Tiedosta voidaan jalostaa valmiiksi johdettuja lukuja, kuten tuottoa. Lisäksi poiminnan yhteydessä voidaan tehdä tiedolle tarkistuksia. Tarkistukset ja jalostaminen kuitenkin tehdään useimmiten ETL:n muokkausvaiheessa. (Hovi ym. 2001, 80.)

Tietojen poiminnan jälkeen tiedot tulee siirtää latausta varten. Siirtoa varten on suunniteltu omia, juuri tiedonsiirtoon suunniteltua ohjelmistoja. Ohjelmistojen avulla tiedon siirtäminen voidaan ajoittaa tai käynnistää automatisoidusti. Siirron valmistuttua tietoa

aletaan muokata. Muokkaus-vaihe jakautuu kahteen erilaiseen tehtävään. Ensimmäinen tehtävä on tarkastaminen. Tarkastuksessa tehdään mm. seuraavanlaisia toimintoja. Duplikaattirivien löytäminen ja niistä raportointi ja poisto. Tyhjien, pakollisiksi määriteltujen, sarakkeiden löytäminen ja raportointi. Lukumäärätarkastus, kuten alkuperäisen aineiston rivit poikkeavat tavanomaisesta. Viite-eheysrikkomusten löytäminen ja raportointi, kuten myyntirivejä joista puuttuu myynti. Muototarkistukset, kuten postinumeron täytyy olla numeerinen ja viisi merkinen tai henkilötunnuksessa tulee olla syntymävuosi sekä henkilötunnuksen loppuosa. Tarkisteet, esimerkiksi henkilötunnuksen toiseksi viimeinen numero tulee olla pariton miehellä ja parillinen naisella. Raja-arvotarkistukset, kuten iän on oltava nolla tai enemmän. Virheellisiksi havatut tiedot voidaan hylätä tai kirjoittaa ne virhelistaan tai -tauluun. Mikäli virheelliset rivit halutaan säilyttää tietovarastossa, voidaan ne merkitä virheelliseksi, etteivät tietovaraston käyttäjät käytä virheellistä tietoa raporteissaan ja analyyseissaan. (Hovi ym. 2009, 56.)

Muokkaus-vaiheen toinen tehtävä on tiedon muuntaminen. Joissakin tapauksissa tiedot voidaan siirtää suoraan sellaisenaan mutta joissakin tilanteissa muuntaminen on tärkeää. Muuntamisen tarkoituksena on muokata operatiivisesta järjestelmästä tullut tieto helpokäyttöiseen ja raportointia tukevaan muotoon. Muokkauksessa tehdään mm. seuraavia toimintoja. Koodin selittäminen käyttäjille, esimerkiksi postinumeron viereen noudetaan postitoimipaikka. Yhdistelmätiedon purku, henkilötunnuksesta voidaan erotella syntymäpäivä, ikä ja sukupuoli omaan sarakkeeseen. Tietojen täydentäminen, kuten opiskelijanumeron perusteella liitetään omat rivit aloitusvuodesta ja opiskelijaryhmästä. Yksikkömuunnoksia, esimerkiksi valuuttojen ja mittojen yhdenmukaistaminen. Tunnuslukujen ja mittareiden laskeminen, kuten asiakaskatteen ja -tuottavuuden laskeminen valmiiksi omaan sarakkeeseen. Summaus, esimerkiksi lasketaan valmiiksi opiskelijaryhmän keskiarvo. Vanhemmissa operatiivisissa järjestelmissä on usein tehty erilaisia virityksiä tiedolle, jotka eivät ole selkokieleisiä, ja eivätkä sovellu ilman muokkaamista tietovarastoon. Tällaisen tiedon korjaaminen ja selvittelyä voidaan kutsua ”siivoamiseksi”. Tätä siivoamista suoritetaan usein muokkausvaiheessa. (Hovi ym. 2009, 56–57.) Siivoamisen ennalta suunnittelu on ongelmallista, sillä on vaikeaa arvioida, kuinka paljon vanhasta operatiivisesta järjestelmästä tulee puhdistettavaa tietoa. Useimmissa tapauksissa tietojen huono laatu paljastuu vasta ensimmäisien latauksien yhteydessä. (Hovi ym. 2001, 82.)

Tiedon muokkaamisen jälkeen tieto on valmiina lataus-vaiheeseen. Latauksen aikana muokatut tiedot ladataan tietovarastoon. Tapahtuma- tai faktatyyppiset rivit lisätään useimmissa tapauksissa suoraan vanhojen, samantyyppisten tietojen, perään. Staattiset ja mitattavat tiedot useimmiten kirjoitetaan vanhojen tietojen päälle. Tällaista tietoa ovat esimerkiksi asiakastiedot. Mikäli jokaisella latauskerralla kirjoitettaisiin uudet rivit samasta asiakastiedosta, olisi tietovarasto täynnä samojen asiakkaiden tietoja. Mikäli kaikki staattiset tiedot halutaan säilyttää, voidaan ne historioida tietovaraston historia-sektorille. Lataaminen tietovarastotietokantaan voidaan tehdä tietokanta-ohjelmistojen sijoita-komennolla, joka sijoittaa tiedon tietovarastotietokannan riveille. Toinen vaihtoehto on käyttää tietokantasovelluksen omia latausohjelmia. Latausohjelmien hyötynä on niiden tehokkuus. Useimmiten tieto ladataan ETL-välineillä, sillä ne pystyvät kirjoittamaan kantaan joko suoraan tai käyttämällä tietokantatuotteiden latausohjelmia. (Hovi ym. 2009, 58.)

ETL-prosessi voidaan ajastaa tapahtumaan haluttuna aikana, useimmiten ajastus on tehty ilta- tai yöaikaan, sillä operatiivisten järjestelmien käyttö on erittäin vähäistä näinä aikoina. Ajastus voidaan tehdä ETL-välineen omilla ajastin-työkaluilla. Mikäli kyseistä työkalua ei haluta käyttää, voidaan ottaa käyttöön käyttöjärjestelmän ajastuspalvelu. ETL-prosessia voidaan seurata ajonseurantaloikeilla. Lokeihin tallennetaan ETL-prosessin läpivientiä kuvaavaa tietoa. Lokeista selviää, esimerkiksi milloin kukin ajo käynnistyi, kuinka kauan ajossa meni aikaa, montako riviä käsiteltiin, minkälaisia virhetilanteita tapahtui jne. Tätä tietoa voidaan käyttää hyväksi virheiden tutkimisessa, latausprosessin kehittämisessä ja seurannassa. (Hovi ym. 2009, 58.)

3.10 Tietovarastoinnin työkaluohjelmistot

Tietovarastointiin liittyvät vahvasti erilaiset työkalut jotka helpottavat tietokannan perustamista ja ylläpitoja. Ilman työkaluja tietovaraston toteutus olisi erittäin vaikeaa. Tietovarastoinnin ohjelmistot voidaan jakaa viiteen pääalueeseen. ETL -välineet, tietovarastotietokantatuotteet, raportointivälineet, metatavälineet ja mallinnusvälineet. Raportointivälineet ovat yksi suurimmista työkaluohjelmistojen ryhmästä, jopa niin isoja, että niille on omistettu kokonaisia kirjoja. Tämän takia ne käsitellään omassa kappaleessa Business Intelligence -nimekkeellä. (Hovi ym. 2009, 60.)

ETL-välineet on suunniteltu latausprosesseiden toteuttamiseen. Useimmissa ETL-työvälineissä on mukana tuki tiedonsiirtoon, graafinen käyttöliittymä tietovirtojen suunnitteluun, mahdollisuus suorittaa tietokantaa komentoja ja tietokantaproseduureja. Tärkeimmät ominaisuudet liittyvät juuri tietojen käsittelyyn. Tärkeää onkin, että ETL-työkalu voidaan kytkeä useisiin tietolähteisiin ja muokkaamaan sekä yhdistelemään näiden sisältämää tietoa ja metatietoa. Muita ominaisuuksia ovat mm. uudelleenkäynnistettävyyys eri työnkulun pisteissä, latauksen ajastaminen sekä jatkuva lataus reaaliaikaiseen tietovarastoa varten, virheiden tutkiminen, raportointi ja niistä hälyttäminen, selkeät lokit ja tilastot latausajoja koskien, versiohallinta. Yksi suosituimmista ETL-välineistä tällä hetkellä on Informatica Corporationin valmistama Informatica-järjestelmä. Kyseinen ohjelmisto on ollut pitkään markkinoilla, ja sillä on vankkumaton tuki käyttäjien keskuudessa. Informatica on lisäksi pysynyt täysin itsenäisenä eikä sitä ole ostettu isojen tietovarastoyritysten toimesta. (Hovi ym. 2009, 60–61.)

Itsenäisten ETL -työkalujen lisäksi tietokantatoimittajilla on omat välineensä ETL-prosessien suunnitteluun ja toteuttamiseen. Esimerkiksi Microsoftin SQL Server-ohjelmistoon kuuluu SQL Server Integration Services-työkalu ja Oraclen tuotteista löytyvä Oracle Warehouse Builder. Nämä vaihtoehdot ovat yleensä erittäin edullisia sillä ne tulevat paketoituna tietokannan mukana. Maksullisten ja suljettujen ohjelmistojen lisäksi avoimen lähdekoodin ratkaisuja on ilmennyt markkinoille. Nämä tuotteet ovat hyvin samantyyppisiä kuin suljetut vastineet, mutta niitä valmistavat organisaatiot eivät välttämättä tarjoa samanlaista tukea kuin maksulliset ratkaisut. Tuotteiden etuna on niiden edullisuus, sillä niitä voidaan ottaa käyttöön täysin maksutta. (Hovi ym. 2009, 60–61.)

Tietovarastojen tietokantoina käytetään usein yleiskäyttöisiä, myös operatiivisissa järjestelmissä käytettäviä SQL-relaatiotietokantoja. Isoimpia tietokantojen tuotteiden valmistajia ovat IBM, Oracle sekä Microsoft omilla tietokannoillaan. Näiden jättiläisten lisäksi on tarjolla pienempi, avoimen lähdekoodin relaatiotietokanta MySQL, jonka tällä hetkellä omistaa Oracle. Nämä tuotteet on suunniteltu operatiivisille sovelluksille, joten ne sisältävät raskasta, ajantasaista tapahtumankäsittelyä tukevia ominaisuuksia. Tällaisia ominaisuuksia ovat mm. lukitukset, toipumisominaisuudet ja online-varmistukset. Lisäksi niihin on lisätty vuosien saatossa uusia tietovarastointiominaisuuksia, kuten tehokkaampi toiminta tietovarastokäyttöön, tietojen pakkaaminen ja tuki summaustauluil-

le. Uusien tietokanta ominaisuuksien lisäksi jotkut tietokantatuotteet sisältävät ETL- ja raporttitoimintoja. Useimmat tietokantavalmistajat ovat paketoineet tuotteitaan laitetointajien kanssa. Tästä paketoinnista käytetään nimeä Data Warehouse Appliance, jossa on valmiiksi muokattu palvelin ja siinä valmis tietokantajärjestelmä. Tämä nopeuttaa käyttöönottoa, sillä asiakas saa valmiin, asiakkaan vaatimuksilleen ja tietomäärille suunnitellun, ratkaisun. (Hovi ym. 2009, 62.)

Joissakin tapauksissa tavalliset relaatiotietokannat eivät sovellu tietovarastojen tietokannaksi, kuten tilanteissa, joissa tietovarasto on erittäin suuri. Tällöin tarvitaan erikoistuotteita, jotka on suunniteltu varta vasten tietovarastokäyttöä varten. Yksi suurten tietovarastojen tietokanta on Teradata jota valmistaa Teradata Corporation. Teradataa käytetään maailman isoimmissa tietovarastoissa, sillä sen erikoisominaisuutena on erittäin vahva rinnakkaisprosessointi. Rinnakkaisprosessoinnissa kytketään rinnan useita, omilla kovalevyillä varustettuja standardipalvelimia. Tällä tavalla työ jakautuu useamman palvelimen kesken. (Hovi ym. 2009, 62.)

Metatietovälineet ovat ongelmallisempi työkalujen osa-alue, sillä lähes kaikkien raportti- ja latausvälineet tukevat metatiedon tallentamista. Ongelmana tässä on se, että metatieto leviää eri paikkoihin. Leviämisen myötä metatiedosta tulee paikallista, mikä ei ole toivottavaa, sillä sen tulisi olla yhteiskäyttöistä, ylläpidettyä ja ajan tasaista. Toisin sanoen, metatieto on sellaista tietoa, minkä tulisi olla keskitetyssä tietovarastossa, eikä hajallaan eri työkaluissa. Tätä ongelmaa vastaan taisteltiin aikaisemmin tietohakemistotuotteilla, jotka pitivät kirjaa tietojen välisistä viittauksista, ja joihin tiedot kuvattiin. Valitettavasti tietohakemistojen käyttö on vähentynyt, osittain huonojen välineiden takia, osittain toimintatapojen muutosten johdosta. Metatietotyökaluja on tullut markkinoilla, mutta ne ovat olleet kalliita ja turhan monimutkaisia ja tästä syystä ne eivät ole olleet kovin suosittuja. Huonon tarjonnan takia yritykset ovat alkaneet käyttää tekniselle metatiedolle ETL-välineiden ominaisuuksia ja rakentavat liiketoiminta-metatiedolle oman tietokantasovelluksen, jotka ovat yleensä erittäin helppokäyttöisiä. (Hovi ym. 2009, 62–63.)

Tietovarastojen mallinnukseen käytetään CASE-välineitä (Computer Aided Software Engineering, Tietokoneavusteista sovellussuunnittelua). Näillä välineillä luodaan tietomalleja ja generoidaan tietokannan luontilauseita. Ne tekevät ohjelmointityöstä helpom-

paa. CASE -välineitä ovat mm. ErWin, Sybase Powerdesigner ja Microsoftin Vision Professional-versio. (Hovi ym. 2009, 63.)

3.11 Business Intelligence

Business Intelligencen (BI) avulla organisaation henkilöstö pääsee hyödyntämään oman organisaation toimintaa kuvaavaa informaatiota. Tätä informaatiota tarvitaan viisaiden päätösten tekemisessä ja toiminnan ohjaamisessa oikeaan suuntaan. BI ilmenee käytännössä useilla eri tavoilla, sitä ovat mm. Exceliin tietoja keräävä kontrolleri, raportteja päivittävä myyntihenkilö tai johtaja, joka tarkastelee johtamansa organisaation tunnuslukuja. Jokainen näistä henkilöistä on loppukäyttäjä BI-ratkaisuissa, joita he voivat hyödyntää tehokkaasti silloin kun he tarvitsevat, tuntematta tietojärjestelmiä tai tietokantojen rakenteita. Parhaimmillaan BI -ratkaisut ovat todella yksinkertaisia, joissa tieto on esitetty visuaalisesti ja helposti, että tiedon hakeminen onnistuu jokaiselta käyttäjältä ilman suurta perehtymistä. (Hovi ym. 2009, 74.)

Nykypäivänä ongelma ei ole tiedon talteenotto, sillä levykapasiteettien kasvaessa ja tallennusmekanismien kehittyessä voidaan melkein kaikki organisaatiolle hyödyllinen tieto tallentaa. BI-ratkaisut tarkoituksena on vastata haasteeseen, jonka suuren tietomäärän jalostaminen, analysoiminen ja tehokas hyödyntäminen asettavat. Parhaissa tapauksissa organisaation hyödynnettäväksi saadaan tietovarastoista ja operatiivisten järjestelmistä informaatiota, joka ilman BI-ratkaisuja olisi jäänyt huomaamatta ja käyttämättä. Tästä syystä BI-ratkaisuihin investoidaan nykyään yhä enemmän yrityksissä ja organisaatioissa. (Hovi ym. 2009, 74.)

Tiedon määrä kasvaa jatkuvasti niin yrityksen sisä- kuin ulkopuolella. Tämä tarkoittaa sitä, että tietoa on ja tulee olemaan enemmän kuin koskaan ennen. Tietoa muodostuu esimerkiksi matkapuhelinoperaattorien tietokantoihin pelkästään siitä, että käyttäjät pitävät matkapuhelintaan päällä. Suomessa suurimmilla operaattoreille on miljoonia asiakkaita, joista jokaisesta muodostuu kyseistä tietoa. Tietoa muodostuu jo päivätasolla todella paljon. BI-ratkaisuilla tätä massiivista tiedon määrää voidaan käsitellä helpommin. (Hovi ym. 2009, 74.)

BI-ratkaisut helpottavat ja nopeuttavat päätöksentekoa, sillä nykypäivän organisaatioissa on entistä vähemmän aikaa tehdä päätöksiä jotka vaikuttavat liiketoimintaan. Tämä ajan puute on johtanut tilanteeseen, jossa raportteja pitää luoda kuukausi-, viikko- ja jopa päivätasolla. BI-ratkaisujen tulee toimia yhä tehokkaammin ja nopeammin. Tämä tarkoittaa nopeampia lataussyklejä tietovarastosta ja automatisoituja raportteja. (Hovi ym. 2009, 76.)

Päätösten tekemisen nopeutuessa myös analyysien tulee olla paljon monipuolisempia. Monipuolisuus taas lisää tiedon monimuotoisuutta, eli analyyseissa on paljon erilaista, eri lähteistä ja erimuotoista tietoa. Tiedon monimuotoisuuden vuoksi tietovarastointiratkaisut ja -hankkeet ovat muuttuneet entistä haastavammiksi. Haastavissa hankkeissa onnistuminen on kiinni oikein valituista integrointimenetelmistä, tietovarastoarkkitehtuurista ja tiedon laadun varmistamisesta. Integroinnin huomioiminen on yksi tärkeimmistä huomioitavista asioista nykyajan yrityskauppojen ja fuusioiden takia. Kahden tai useamman yrityksen tietojen, kuten asiakasrekistereiden ja myyntijärjestelmien, yhdistäminen on huomioitu useissa toiminnohjausjärjestelmiä toimittavissa yrityksissä. Nykyisissä toiminnohjausjärjestelmissä BI on noussut yhdeksi merkittäväksi kehityskohteeksi. Viranomaisten vaatimukset lisäävät tarvetta hankkia BI-ratkaisuja. Nämä vaatimukset ovat kasvaneet vuosi vuodelta erilaisten kirjanpitoväärinkäytösten, pankkien höveliään lainannon sekä yritysmaailmassa tapahtuneiden skandaali- ja korruptiotapausten myötä. BI-ratkaisut pyrkivät vastamaan näihin asetettuihin vaatimuksiin. (Hovi ym. 2009, 76.)

Jokaisessa BI -ratkaisussa on tärkeää, että noudatetaan seuraavia ominaisuuksia. Ilman näitä ominaisuuksia BI ei palvele organisaatiota, vaan päätökset joudutaan tekemään ilman faktatietoa. Ensimmäinen näistä on analytytisyys. BI:n tarkoituksena on hyödyntää informaatiota mahdollisimman analytytisesti. BI-ratkaisut hyödyntävät operatiivisiin järjestelmiin kerättyä dataa, josta se muodostaa uutta informaatiota organisaation toiminnan tehostamiseksi. (Hovi ym. 2009, 82.)

Toisena ominaisuutena on kvantitatiivisuus, eli määrällisyys. BI -ratkaisujen käsittelemä tieto on pääosin numeerista faktatietoa, kuten myyntilukuja, varaston kiertonopeuksia yms. Näitä määrällistä lukuja käsitellään tietojen eri ulottuvuuksien, kuten aika ja paikka, avulla niin, että luvista muodostuu ihmiselle helposti käsiteltävä kokonaisuus.

BI-ratkaisut antavat käyttäjilleen vastauksia kysymyksiin, kuten oppilaiden valmistusmäärä kymmenen viime vuoden aikana tietojenkäsittelynkoulutusohjelman peliohjelmoinnin suuntauksesta. (Hovi ym. 2009, 82.)

Kolmantena ominaisuutena on liiketoimintalähtöisyys. BI:llä tarkoitetaan älykästä liiketoiminnan johtamista. Valitettavasti tämä jää useimmiten vähimmälle huomiolle kun BI-ratkaisuja kehitetään pelkästään teknologisista lähtökohdista. Tällaisessa tilanteessa on useimmiten hankittu monipuolisia ohjelmistoja, joilla jokainen organisaation ongelmat ratkaistaan näennäisen helposti, mutta liiketoiminnan tarpeita ei ole otettu ollenkaan huomioon. Tästä taas seuraa usein se, että IT-osasto valmistaa ratkaisun ja toivoo, että käyttäjät löytävät kaikki tarvittavat omiin tarpeisiinsa. Useimmiten käyttäjät kuitenkin eivät löydä haluamaansa ja hylkäävät IT-osaston rakentaman ratkaisun. BI-projekteissa tulisikin olla edustettuna IT- ja businesspuoli. Tällöin kuvattu tilanne voidaan välttää eikä organisaatiolle tule tarpeettomia hankintoja ja IT -puolelle tarpeettomia BI -ratkaisujen rakentamisia. (Hovi ym. 2009, 82.)

Viimeisenä ominaisuutena on koko organisaatio huomioon ottaminen. Moni saattaa kuvitella, että ainoat henkilöt organisaatioissa, jotka tarvitsevat tunnuslukuja ja parempaa informaatiota ovat johtajat. Nykypäivä tämä ajatusmaailma ei kuitenkaan ole oikea. Useimmissa tapauksissa BI-ratkaisujen käyttäjät koostuvat kolmesta ryhmästä. Nämä kolme ryhmää ovat johto, tehokäyttäjät sekä kontrollerit ja tiedon kuluttajat. Johto on itse asiassa pienin ryhmä joka tarvitsee BI:n tuottamaa tietoa ja suurin osa käyttäjistä on organisaation työntekijöitä. Tästä voidaankin päätellä, että tietoa tarvitaan joka puolella organisaatiota, eikä perinteinen näkemys siitä, että johtoporras tietää kaiken ja alaiset vain tarvittavan päde nykypäivänä. (Hovi ym. 2009, 82.)

4 ISAT-hankkeen tietovarasto

4.1 Alkutilanne

ISAT:n oppilaitoksilla on tällä hetkellä samanlainen tilanne, kuten on kuvattu luvussa 3.2. Käytössä on useita erilaisia operatiivisia järjestelmiä, jotka tallentavat tietonsa

omiin tietokantoihinsa. Tietokantoihin tallennettuihin tietoon päästään käsiksi tietokantaa käyttävän operatiivisen järjestelmän kautta. Tietoa ei kerätä tällä hetkellä keskitettyyn tietovarastoon, josta käyttäjä voisivat käsitellä ja jalostaa kaikkea tarvitsemaansa tietoa halutulla tavalla. Tietoa ei siis saada niin helposti saatavilla kuin olisi toivottavaa. Hovi, Hervonen ja Koistinen (2009, XI) nimittävät tätä tilannetta siiloutumiseksi, tietoa on paljon mutta se on saatavilla erilaisista lähteistä.

Siiloutumisen takia ammattikorkeakoulujen eri organisaatioilla saattaa olla erittäin erilainen kuva oppilaitoksen toiminnasta. Näiden seikkojen takia joudutaan usein tilanteeseen, jossa koko oppilaitosta käsittelevää raportointia varten joudutaan keräämään tieto jokaisesta ”tietosiilosta” erikseen. Usein tieto joudutaan keräämään useamman henkilön toimesta, tehden kokonaisraportoinnista on erittäin hidasta ja kömpelöä. Tämä ongelma on kuvattu luvussa 3.2.1, nykyisten järjestelmien raportit. Käytössä on valmiiksi ohjelmoituja raportteja jotka on ennalta ohjelmoitu järjestelmän toimittajan toimesta. Tämä tarkoittaa sitä, että tietoa ei saada käsitellä ja valikoida vapaasti omanlaisiin tarpeisiin vaan joudutaan ajamaan raportteja jotka tuovat ennalta määritetyt tiedot näkyville. Mikäli tietoa halutaan jatkojalostaa, joudutaan ne syöttämään taulukkolaskimeen, kuten esimerkiksi Excel-järjestelmään. (Savonia 2011.)

Tietojen siirto manuaaliryönnä taulukkolaskentaohjelmasta lisää virheitä mikä laskee raporttien luotettavuutta. Lisäksi taulukkolaskinten käyttö on erittäin aikaa ja resursseja vievää, sillä tietojen hajanaisuuden ja erilaisten operatiivisten järjestelmien takia tietoa joudutaan muokkaamaan syöttövaiheessa tiedonsyöttäjän toimesta. Näihin muutoksiin kuuluu muun muassa käytettävien koodien yhtenäistäminen. Näiden syiden takia ammattikorkeakoulujen johdolla ei välttämättä ole aina saatavilla ajankohtaisinta ja ennen kaikkea oikeaa tietoa johtamista varten. (Savonia 2011.)

Siiloutumisesta huolimatta tiedon hajanaisuus ei ole välttämättä niin suuri ongelma kuin palveluita ja osaamistaan tarjoilevilla yrityksillä, joilla on paljon asiakkaita. Tällaisissa yrityksissä on suuri tarve tehdä erilaisia arvioita asiakkaan arvosta yritykselle sekä tehdä asiakasprofilointia. Tämä auttaa saamaan asiakkaasta paremman kokonaiskuvan jolloin asiakkaalle on helpompi tarjota hänen tarvitsemiaan palveluita. Tällaisille toiminnoille ammattikorkeakouluissa ei ole suuri tarve. Tämä ei silti tarkoita sitä, ettei ammattikorkeakouluilla olisi tarvetta tietovarastointiratkaisuille. Kokonaisraportoinnin automati-

soiminen, helpottaminen ja tehostaminen ovat varmasti erittäin tervetulleita muutoksia suurille ammattikorkeakouluille. Tiedon määrä, joka opiskelijahallinnosta, henkilöstöhallinnosta sekä taloushallinnosta muodostuu vuosittain, on valtaisa ja tämän massiivisen tietomäärän tehokkaampi hyödyntäminen tuo huomattavasti lisäarvoa kummallekin ammattikorkeakoululle. (Savonia 2011.)

4.2 Esitutkimus

ISAT-hankkeen tietovaraston rakentaminen aloitettiin alkuvuodesta 2011. Ensimmäisenä vaiheena tietovaraston rakentamisessa oli esitutkimus, joka ISAT-hankkeen tietovarastossa toteutettiin projektisuunnitelman yhteydessä. Projektisuunnitelmassa selvitettiin seuraavat projektin liittyvät kohdat: yleiskuvaus, tavoitteet, riskit, hyödyt, raja-
aus, arkkitehtuurin linjaus, kustannukset ja projektin vastuuhenkilöt. (Savonia 2011.)

Esitutkimuksen ja projektisuunnitelman ero on niiden karkeudessa ja siitä, mitä ne käsittelevät. Esitutkimuksessa on tarkoitus luoda näkemys siitä, minkälainen projekti on kyseessä ja kuinka se suoritetaan. Lisäksi se ottaa kantaa onnistumismahdollisuuksiin ja siihen, onko projekti taloudellisesti kannattava. Esitutkimus ei yleensä mene syvälle projektin eri osa-alueisiin, eli se on yksityiskohdiltaan hyvin karkea. Esitutkimuksen valmistuttua siitä syntyvä raportti hyväksytetään projektista vastaavilla henkilöillä, tästä lähtien projektiryhmä, jolloin projektin toteuttaminen voidaan aloittaa. Mikäli esitutkimus ei tyydytä projektiryhmää, tulee tehdä lisäselvitys. Esitutkimuksen hyväksymisen jälkeen aloitetaan projektisuunnitelman toteuttaminen. Projektisuunnitelmassa listataan kaikki tehtävät mitkä projektin aikana on suoritettava. Näille tehtäville asetetaan myös aikataulu. Projektisuunnitelma, toisin kuin esitutkimus, on hyvin yksityiskohtainen, eli sen on tarkoitus ottaa huomioon kaikki projektissa tapahtuvat asiat. Projektisuunnitelma on hieman kuin projektin kartta, jota seuraamalla pääsee haluamaansa kohteeseen, eli tässä tapauksessa toimivan tietovaraston toteuttamiseen. (Hovi ym. 2009, 138–140.) Jakoa esitutkimuksen ja projektisuunnitelman välillä ei tehty ISAT-hankkeen tietovarastoprojektissa. Tästä huolimatta projektisuunnitelmassa otettiin kantaa hyvin tarkasti samoihin esitutkimuksen osa-alueisiin, joihin ammattilaiset, kuten Hovi ym. (2009, 138–139) suosittelevat ottamaan huomioon tietovarastoprojektia aloittaessa.

Vaikka projektisuunnitelmassa sivutaan samoja asioita kuin esitutkimuksessa, ammattilaiset suosittelevat tekemään esitutkimuksen aina ennen projektisuunnitelman tekoa. Tämä johtuu siitä, että esitutkimuksessa selvitetään, onko tietovarasto järkevää ja kannattavaa tehdä. Mikäli jo esitutkimuksen aikana huomataan, että käynnistymässä oleva tietovarastoprojekti ei ole yritykselle kannattava tai jopa mahdoton toteuttaa, voidaan se lopettaa ennen suurempia kuluja. Vanha sanonta: ”tutkitaan, ennen kuin hutkitaan” pätee hyvin tietovarasto projekteissa, joissa on mahdollista menettää paljon resursseja turhaan kun huomataankin, että projektia ei voida toteuttaa toivotulla tavalla. (Hovi ym. 2009, 138)

4.3 Projektisuunnitelma

4.3.1 Johdanto

ISAT-hankkeen tietovaraston projektisuunnitelma alkaa johdannolla, joka selvittää, mistä ISAT-tietovarastosta on kyse. Johdannossa on tarkoituksena kuvata projektin aloittamiseen johtaneet syyt ja yrityksen nykytilanne (Pelin 2008, 88). ISAT-hankkeen tietovarastoprojektin projektisuunnitelman johdannon mukaan ISAT-hankkeen tietovarastoprojektin tarkoituksena on tehostaa ja palvella Savonian ja PKAMK:n omia tilastointi- ja mittarointitarpeita. Lisäksi tietovaraston ansiosta kummatkin oppilaitokset voivat tehdä paremmin vertailua sekä mittarointia. Projektin tavoitteena on luoda tilanne, jossa operatiivisista järjestelmiin tuleva tieto siirretään automaattisesti tietovarastoon, jossa se voidaan jalostaa halutulla tavalla raportointia ja mittarointia varten. (Savonia 2011.)

Oppilaitoksen omien mittareiden ja raporttien lisäksi tietovarastoon kerättyä tietoa siirretään valtakunnalliseen tietokantaan, josta Opetus- ja kulttuuriministeriö ja tilastokeskus saavat korkeakouluja koskevaa tilastotietoa (Savonia 2011). Tämän tietovaraston nimi on Rakenteellisen Kehittämisen Tukena Tietohallinto-XDW, tästä lähtien Raketti-XDW. Raketti-XDW:n tarkoituksena on toteuttaa käsitemalli, sekä tähän käsitemalliin perustuva tietovarastopalvelu opetusministeriön, ammattikorkeakoulujen ja yliopistojen käytettäväksi. Käsitemallin tarkoituksena on yhtenäistää korkeakoulujen raportointia, jolloin korkeakoulujen tuottama tieto on yhdenvertaista ja vertailukelpoista, eli sitä voi-

daan verrata keskenään puolueettomasti. Vertailukelpoisuuden myötä korkeakoululaitosten päätösten teko helpottuu. Raketti-XDW:n yhteisen tietovaraston tarkoituksena on yksinkertaistaa ja tehostaa viranomaisraportointia. (CSC 2012.)

4.3.2 Hyödyt

Johdannon jälkeen siirrytään projektin hyötyihin. Hyödyt on tärkeää kirjata projektisuunnitelmaan, sillä on tärkeää, että kirjattuja hyötyä pystytään vertaamaan aiheutuviin kustannuksiin (Sininen Meteoriitti 2012). Hyödyiksi ISAT-hankkeen tietovarastoprojektissa mainitaan tilastointirutiinien, perusjärjestelmien tietojen yhtenäistäminen ja tietojen kirjauskäytäntöjä. Näiden hyötyjen ansiosta tietojen ajantasaisuus ja oikeellisuus paranee. Hyödyissä mainitaan myös luvussa 4.1, alkutilanne, mainittu Excel-tilastointi. Tietovaraston ansiosta Excel-tilastointia ja -mittarointia ei tarvita, vaan tietovarasto suorittaa kaikki tarvittavat laskennat automaattisesti. Viimeisenä otetaan myös kantaa johdon kaipaamaan reaaliaikaiseen tietoon, sillä tietovarasto parantaa tiedontuotannon laatua, joka tuottaa täsmällisempää informaatiota ammattikorkeakoulujen johdolle. (Savonia 2011.)

4.3.3 Vaiheet

Projektisuunnitelman seuraavassa luvussa on asetettu projektista toteutettavat tehtävät vaiheisiin. Projektin vaiheistamisessa on tarkoituksena jaksottaa projektissa toteutettavat tehtävät ajallisesti peräkkäisiin vaiheisiin. Päämääränä vaiheistamisessa on päätöksenteon helpottaminen, sillä kun projektissa toteutettavat tehtävät on vaiheistettu, voidaan vaiheen lopussa tehdä jatkosuunnitelmia projektin seuraaviin vaiheisiin. Lisäksi jokaisen vaiheen lopussa syntyy tulos, joka on mitattavissa, kuten prototyyppi, projekti-dokumentti tai selvitys jne. Vaiheiden ansiosta projektin johdon ei tarvitse sitoutua päätöksiin, jotka tuntuvat epävarmoilta. Vaiheiden välejä voidaan kutsua tarkistuspisteiksi, joista johtoryhmä näkee mihin projekti on etenemässä ja mikä eri vaiheiden lopputulos on. (Pelin 2008, 99–100.)

ISAT-projektin tietovaraston projektisuunnitelmassa projekti on vaiheistettu seuraavalla tavalla: sisällölliset määrittelyt, tekniset määrittelyt, perustietojärjestelmien valmistelu, tekninen toteutus, testaus ja käyttöönotto. Sisällöllinen määrittely sisältää projektin hen-

kilöstö tarpeiden määrittelyn, eli tilastointi-, laatu- ja johtohenkilöstön nimeäminen. (Savonia 2011.)

Tekninen määrittely sisältää teknisen alustan valinnan eli mitä tietokantapalvelinta käytetään, raportointivälineiden valinta ja BI-välineiden valinta. Liittymät perusjärjestelmistä tietovarastokantaa eli kuinka tieto siirretään oppilaitoksen operatiivisista järjestelmistä tietovarastoon. Tietovaraston taulukuvaukset, eli käytetäänkö Raketti-XDW:n tietorakenteita vai luodaanko muunneltuna versio oppilaitoksen omien operatiivisten järjestelmien tietorakenteista. Tiedonsiirtoliittymä Raketti-XDW:hen, eli kuinka tieto siirretään oppilaitosten omasta tietovarastosta Raketti-tietokantaan. (Savonia 2011.)

Perusjärjestelmien valmistelussa oppilaitosten operatiiviset järjestelmät valmistellaan tiedon siirtämiseen järjestelmistä tietovarastoon. Teknisessä toteutuksessa suunniteltu tietovaraston eri osa-alueet kuten tietokanta, operatiivisten järjestelmien ja tietovaraston liittymät, tietojen jalostus ja raportointi toteutetaan ja valmistetaan käyttöönottoa varten. Teknisen toteutuksen tulokset testataan testausvaiheessa. Testauksen jälkeen viimeisenä vaiheena on käyttöönotto. Käyttöönotto tapahtuu, kunhan jokainen tilastointi- ja mittarisuus on valmistunut. (Savonia 2011.)

Vaiheittainen suunnittelu ja toteuttaminen takaavat sen, että samoja virheitä ei toisteta jokaista osa-aluetta suunnitellessa ja toteutettaessa. Vaiheittain toimiessa tietovaraston suunnittelusta ja toteutuksesta vastaavat henkilöt oppivat uusia toimintatapoja sekä -menetelmiä ja ennen kaikkea saavat tärkeää kokemusta tietovarastoprojekteista. Näin henkilöistä, jotka eivät olleet ennen toteuttaneet tietovarastoprojekteja oppivat tietovarastoprojekteihin liittyvät ”niksit” jo projektin ensimmäisten vaiheiden aikana. Tällä tavoin tietovaraston jatkokehittäminen ja seuraavien vaiheiden toteuttaminen helpottuu. Kehittämisestä vastaavien henkilöiden tietotaidon kehittymisen lisäksi saadaan tärkeätä tietoa siitä, toimivatko välineet, menetelmät, tekniset ratkaisut ja laitteet suunnitellulla tavalla ensimmäisessä vaiheessa. Ensimmäistä vaihetta voidaankin pitää testivaiheena, sillä siinä nähdään, mikäli jokin osa-alue ei toimi. Mikäli kaikki osa-alueet toimivat, voidaan seuraavat vaiheet suorittaa paljon nopeammin. Tietovaraston vaiheittaista rakentamista kutsutaan myös inkrementaaliseksi rakentamiseksi. (Hovi ym. 2009, 132.)

4.3.4 Organisaatio

Projektin toteuttamiseen tarvitaan aina henkilöitä, jotka pystyvät suoriutumaan projektissa vaadittavista tehtävistä. Projektin onnistumisen kannalta olennaista on, että projektin jokaiselle henkilölle annetaan rooli hoidettavaksi. Roolin myötä henkilö saa omat vastualueen tai vastualueet joita hänen tulee hoitaa. Rooleja jakaessa on tärkeää pitää mielessä, että henkilölle asetettu rooli on hänen kykyjensä mukainen. Tämä tarkoittaa sitä, että palaverimuistioita kirjoittavaa henkilölle ei aseteta vastuualueeksi ohjelmointia tai tietovaraston käyttöönoton valvontaa. Tämä ei silti tarkoita, että yhdellä henkilöllä olisi ainoastaan yksi rooli, vaan rooleja voi olla useampia, kunhan henkilölle asetetut roolit sopivat hänen kyvyilleen. Kaikkia projektissa esiintyviä rooleja ei aina voida tai tarvitse asettaa oman yrityksen väelle, sillä yrityksen ulkopuolista osaamista tarvitaan esimerkiksi operatiivisten järjestelmien toimittajilta ja konsultteja. (Hovi ym. 2009, 151–153.)

ISAT-hankkeen tietovaraston projektisuunnitelmassa on asetettuna ainoastaan projektipäällikön ja tilastojen, mittareiden ja tiedon oikeellisuudesta vastaavan vastuuhenkilön roolit. Muille vastuualueille ei ollut asetettuna henkilöä. (Savonia 2011.) Projektisuunnitelmaan ei ollut siis asetettu kovin montaa roolia, joten monta pakollista roolia puuttui. Hovi ym. (2009, 151) pitävät seuraavia rooleja erittäin tärkeinä tietovarastoprojekteissa: projektin asettaja, projektin ohjausryhmä, johon kuuluu omistaja ja muut jäsenet, projektiryhmä, johon kuuluu projektipäällikkö ja muut jäsenet, toiminnan asiantuntijat, eli tietovastaavat ja käyttäjät, tietovarastoarkkitehti, tietovaraston hoitaja ja käyttöoikeuksien antaja, operatiivisten järjestelmien asiantuntija, ETL-asiantuntija, raportoinnin asiantuntija, tietovarasto-asiantuntija ja tietoturva-asiantuntija. Näiden projektihenkilöiden lisäksi saatetaan tarvita muita asiantuntijoita, kuten konsultteja ja tietoliikenne-asiantuntijoita, riippuen tarpeesta.

Näistä rooleista projektin johtoon kuuluvat projektin asettaja, projektin ohjausryhmä ja projektipäällikkö. Heidän vastuulla on projektin onnistuminen, ja heidän kuulukin tehdä projektia koskevia päätöksiä ja poistaa projektin aikana esiintyviä ongelmia. Projektin johdon rooleista projektin asettajan tehtävänä on aloittaa projekti eli hän on henkilö, jonka päätöksellä projekti aloitetaan. Tavallisesti projektin aloittajana on toimitusjohtaja. Projektin ohjausryhmän tehtävänä on ohjata ja seurata projektia. Heidän tärkein tehtävänsä varmistaa, että projektihenkilöstöllä on aikaa ja mahdollisuudet suoriutua heille

asetetuista tehtävistä. Projektipäällikkö vastaa projektin johtamisesta, joten häntä voidaan kutsua projektin ”toimitusjohtajaksi”, sillä hänen tulee ohjata projektia samalla tavalla kuin yrityksessä toimitusjohtaja ohjaa johtamaansa yritystä. Hänen tärkein tehtävä on hallita ja ohjata projektia projektisuunnitelman mukaisesti. Projektipäällikön muita tehtäviä on mm. seurata kustannusten ja hyötyjen toteutumista, koordinoita työtä sekä toimii laadun valvojana. (Hovi ym. 2009, 152–153.)

Projektin johdon ulkopuolella olevat roolit eivät ole ryhmitetty vaan näitä rooleja kutsutaan muiksi osanottajiksi. Projektiryhmän tehtävänä on tehdä sille määritetyt tehtävät. Projektiryhmässä tulee olla mahdollisimman laajasti osaamista niillä tietovaraston osa-alueilla, jotka projektisuunnitelmassa on suunniteltu toteutettavaksi. Projektiryhmän jäsenet määrittelevät vaatimukset tietovarastoon, tarkistavat tiedon laadun, määrittävät raportit ja kyselyt sekä määrittelevät tietojen sisällön. Tietovarastoarkkitehti vastaa siitä, että tietovarasto noudattaa standardeja ja palvelee toimintaa mahdollisimman hyvin. Hän on tietovaraston ylin valvoja ja hänen vastuullaan on tietovaraston kokonaisuus. Tietovaraston hoitajan tehtävänä on tarkkailla ja hienosäätää tietovarastoa jatkuvasti, jotta sen suorituskyky ei laske. Hän siis ylläpitää tietovarastoa. Lisäksi hoitajan suunnittelee tietovaraston teknisen puolen, eli käytettävän tekniikan. Käyttöoikeuksien antaja määrittää käyttäjille käyttöoikeudet tietovarastosta saataville tiedoille. Hänen tehtävänsä on järjestellä käyttäjät ryhmiin niin, että tietyssä ryhmässä olevat käyttäjät saavat käyttää vain heidän ryhmälleen tarkoitettua tietoa. (Hovi ym. 2009, 152–155.)

Tietovarastoprojektissa tarvitaan lisäksi paljon eri alojen asiantuntijoita. Toiminnan asiantuntijan tuovat omien alueiden, kuten johtajien ja pääkäyttäjien, tiedot ja tarpeet esille. Jokaiselle näille osa-alueelle nimetään oma asiantuntija, eli tietovastaava. Tietovastaavan rooli on erittäin haastava, sillä hänen tehtäviinsä kuuluu mm. varmistaa, että tietovaraston tiedot ja sinne tuotavien tietojen käsittely- ja muodostussäännöt ovat oikein. Lisäksi tietovastaava osallistuu raportoinnin määrittämiin ja tekee tarkistuksia toimitettuihin lähtöaineistoin. Operatiivisten järjestelmien asiantuntijoiden tuottavat tietovarastoon tietoa yrityksessä käytettävistä operatiivisista järjestelmistä. He luovat tietojen poimintaohjelmat, sillä he tuntevat operatiiviset järjestelmät parhaiten. Tarvittaessa operatiivisten järjestelmien asiantuntijat myös yhdistelevät ja summaavat tietoa. Tämä rooli on tietovarastoprojekteissa erittäin tärkeää, sillä tietojen poimintaa saattaa mennä todella paljon aikaa.

ETL-asiantuntijan tehtävä on hakea, muokata ja ladata tietovarastoon operatiivisten järjestelmien tiedot ETL-latausohjelmilla tai perinteisellä ohjelmoinnilla. Tämä rooli on erittäin vaativa ja resursseja vievä, vaikka tietovarasto olisikin pieni. Tästä syystä ETL-asiantuntija tulee tuntea tehtävänsä perinpohjaisesti. Raportoinnin asiantuntija tekee raportit, joita tietovarastossa tarvitaan, eli hänen täytyy tuntea raportointivälineet erittäin hyvin. Tietovarasto-asiantuntija suunnittelee projektissa käytettävän tietovarastoratkaisun ja hallitsee tietovaraston mallintamisen eri tasoilla. Hänen tehtäviinsä kuuluu tietovaraston koon ja sen kasvun suunnittelu. Viimeinen vaadittavista asiantuntijoista on tietoturva-asiantuntija, hän vastaa siitä, että tietovarastossa suunnittelussa ja toteutuksessa noudatetaan yrityksen asettamia tietoturvastandardeja. (Hovi ym. 2009, 152–155.)

Ulkopuoliset konsulttien tehtävänä on auttaa yritystä silloin, kun yrityksestä ei löydy tarvittavaa osaamista. Näitä konsultteja kannattakin käyttää aina, mikäli tällainen puute ilmenee. Ulkopuolisten konsulttien käyttäminen on järkevää, sillä he tuntevat tietovarastoissa käytettävät välineet ja lisäksi heillä on kokemusta erilaisista tietovarastoprojekteista. Tämä onkin johtanut tilanteeseen, jossa useat nykyisistä tietovarastoprojekteissa tehdään ulkoisten konsulttien toimesta. Tämä säästää aikaa ja yrityksen henkilöt voivat keskittyä omien töittensä ja rooliensa hoitamiseen tietovarastoprojektissa. (Hovi ym. 2009, 155.)

4.3.5 Tavoitteet

Oikeanlaisten tavoitteiden määrittäminen on tärkeää, sillä ainoastaan konkreettiset ja realistiset tavoitteet vievät projektia eteenpäin. Ammattilaiset suosittelevat, että ensimmäisestä tietovarastoa rakentaessa osa-alueiden rajaaminen, varsinkin potentiaalisten kohteiden osalta, on elintärkeää hankkeen onnistumisen kannalta. Hyvät rajaukset auttavat myös tavoitteiden asettamista. (Hovi ym. 2009, 138–144.) ISAT-hankkeen tietovarastossa asetettiin oheiset tavoitteet seuraavassa järjestyksessä: tietovarastoon ladataan ensimmäiseksi pelkästään opintohallinnon tietoja ja vasta myöhemmin keskitytään talous- ja henkilöstöhallinnon tietojen siirtämiseen tietovarastoon. Viimeisessä vaiheessa suunnitellaan ja toteutetaan muiden käytettävien järjestelmien, kuten työaikasuunnittelun, tietojen siirtäminen tietovarastoon. (Savonia 2011.)

4.3.6 Riskien hallinta

Riskien hallinta on erittäin tärkeä osa-alue kaikissa projekteissa ja tässä tietovarastoprojekti ei ole poikkeus, sillä projekteihin liittyy aina erikokoisia ja erilaisia riskejä. Riskien hallinnan tärkeydestä kertoo se, että puolet aloitetuista tietovarastoprojekteista Yhdysvalloissa epäonnistuu tunnistamattomien riskien takia. Riskien hallinnan tavoitteena onkin tunnistaa kaikki projektiin vaikuttavat riskit, sillä vain tunnettuihin riskeihin voidaan varautua ennakkoon. Projektia uhkaavat riskit tunnistetaan tekemällä riskikartoitus. (Hovi ym. 2009, 158.)

Riskikartoitus tehdään usein uutta projektia valmisteltaessa. Tämä kartoitus ”elää” projektin alussa, eli kartoitukseen tehdään täydennyksiä, mikäli uusia riskejä ilmenee. Hyvä tapa vielä tunnistamattomien riskien löytämiseen on kysyä tietovarastoprojektiin osallistuvilta henkilöiltä mahdollisia ongelmakohtia projektia aloittaessa. Riskien tunnistuksen jälkeen riskit ryhmitellään ja asetetaan järjestykseen niiden merkitysten mukaisesti. Useasti riskit lajitellaan niiden todennäköisyyden mukaisesti, eli ensin ovat riskit jotka ovat todennäköisiä ja viimeisenä riskit jotka ovat erittäin epätodennäköisiä. Riskien listaamisen jälkeen niiden toteutumisen varalle luodaan toimenpidesuunnitelma, jonka avulla riskin toteutuessa pystytään toimimaan niin, että riskin vaikutukset pystytään minimoimaan. Toimenpidesuunnitelmassa jokaiselle riskille nimitetään vastuuhenkilö. Vastuuhenkilön tehtävänä on ehkäistä määrätyn riskin syntyminen. Riskien toteutumista tarkkaillaan koko projektin ajan, esimerkiksi projektiryhmän kokouksissa. (Hovi ym. 2009, 158.)

ISAT-projektin projektisuunnitelman riskikartoituksessa tunnistettuja riskejä ei ollut listattu kovinkaan montaa, mutta niiden toteutumisen varalla oli luotu tarvittavat toimenpiteet, jotta riskien toteutumiseen ollaan valmiina ja niiden vaikutukset saadaan minimoitua. Ensimmäisenä riskinä pidettiin koko projektin toteutumisen epäonnistumista. Tätä riskiä pidetään erittäin pienenä, sillä projektin epäonnistuessa Savonian ja PKAMK:n operatiiviset järjestelmät toimivat yhä normaalisti. Tämä johtuu siitä, että järjestelmien toimintaan ei tietovarastoa toteuttaessa tarvitse tehdä muutoksia, sillä tietovarastoon ladataan tietoa oppilaitosten operatiivisten järjestelmien tietokannoista ETL-arkkitehtuuria käyttäen, kuten luvussa 3.9.1 on kuvattu. Lisäksi tilastotietoa saadaan operatiivisten järjestelmien raporteista koko tietovarastoprojektin ajan, joten tietoa

saadaan käytettäväksi, vaikka projekti ei toteutuisi aikataulussa tai se epäonnistuisi kokonaan. (Savonia 2011.)

Seuraavana riskinä pidettiin aikataulun venymistä. Ensimmäisenä syynä aikataulun venymiselle pidettiin projektiin osallistuvien henkilöiden työskentely muissa projekteissa. (Savonia 2011.) Hovi ym. (2009, 160–161) ovatkin merkinneet tietovarastoprojekteissa työskentelevien henkilöiden sitoutumattomuuden yhdeksi riskitekijäksi tietovarastoprojekteissa. Heidän mukaansa henkilöiden muita tehtäviä ja vastuita tulisi vähentää ennen tietovarastoprojektin aloittamista. Nyt on luotu tilanne, jossa tietovarastoprojekti on lisätty kaikkien muiden töiden lisäksi henkilöiden tehtäväksi. Tämä saattaa aiheuttaa tilanteen, jossa projektissa työskentelevät henkilöt saattavat jättää tietovarastoprojektin taka-alalle, jolloin projektin eteneminen saattaa hidastua.

Toisena riskinä projektiaikataulun venymiselle pidettiin operatiivisten järjestelmien ohjelmistotoimittajien hitautta toimittaa tarvittavat liitokset tiedon siirtämiseen operatiivisista järjestelmistä tietovarastoon. Tätä riskiä varten ei ollut luotu minkäänlaista toimitasuunnitelmaa. (Savonia 2011.) Tästä huolimatta on tärkeää, että aikataulun venymiseen on varauduttu, sillä ensimmäisessä tietovarastoprojektissa mahdollisiin venymiin on tärkeää varautua. Tästä syystä aikataulusta ei tule tehdä liian tiukkaa, vaan varata reilusti aikaa tietovarastoprojektin toteuttamiseen. Liian tiukkaa aikataulu ei myöskään välttämättä ota huomioon mahdollisia puutteita henkilöstöresursseissa. Henkilöstöresurssin puute voi olla esimerkiksi työntekijöiden kiireet muissa projekteissa ja työtehtävissä sekä mahdolliset poissaolot. (Hovi ym. 2009, 158–161.)

Kolmantena ja viimeisenä riskinä projektisuunnitelmaan oli kirjattu projektiin osallistuvien henkilöiden kokemattomuus ja tarvittavan osaamisen puute. Tämän riskin hallintaan projektihenkilöille järjestetään koulutusta ja teknistä tukea, jonka tarkoituksena on kohottaa projektiin osallistuvien henkilöiden osaamisen tasoa. (Savonia 2011.) Tämän riskin varalla onkin tärkeää tehdä varasuunnitelma, sillä kokemattomuus heijastaa myös projektisuunnitelmassa toisena mainittuun riskiin eli aikataulutukseen. Projektin aikatauluun on tärkeää varata aikaa henkilöstön osaamisen tason tunnistamiseen sekä mahdolliseen lisäkoulutukseen, sillä liiallisen kiireen takia puutteita osaamisessa ei välttämättä huomata ennen kuin on liian myöhäistä. Tällaisissa tilanteissa aikataulu saattaa

venyä turhankin pitkäksi, joka uhkaa projektin valmistumista. (Hovi ym. 2009, 158–161.)

Tietovaraston toteuttamisessa auttavat Raketti-XDW:ssä tuotetut käsittemallit ja kuvaukset, joiden avulla ISAT-hankkeen tietovarasto voidaan yhtenäistään Raketti-XDW:n tietovaraston kanssa. Raketti-XDW:n ansiosta pyörää ei tarvitse keksiä uudelleen vaan voidaan käyttää jo valmiiksi testattua ja yhtenäistä mallia. Tämä säästää aikaa ja pienentää projektin epäonnistumisen mahdollisuutta, sillä projektiin osallistuvien henkilöiden ei tarvitse luoda uutta mallia, jolloin saatetaan kohdata odottamia yllätyksiä. (Savonia 2011.)

4.3.7 Aikataulu

Projektisuunnitelman seuraavassa luvussa käsiteltiin jo riskien yhteydessä mainittua aihetta aikataulu. Aikataulu oli esitetty projektisuunnitelmassa hyvin karkeasti, sillä aikatauluun oli merkattu ainoastaan neljä päätavoitetta johon projektissa pyritään. Jokaiselle päätavoitteelle oli asetettu aika milloin kyseisen tavoitteen tulisi olla valmis. ISAT-tietovarastoprojektin päätavoitteet olivat seuraavat: määrittely ja perustietojärjestelmien kunnostus, joka suunniteltiin aloitettavaksi keväällä 2011. Tekninen määrittely, joka oli ajoitettu myös keväällä 2011. Projektihenkilöiden koulutus, tämäkin päätavoite oli suunniteltu vuoden 2011 keväälle, mutta tämän tavoitteen aikataulusta ei ollut vielä täyttä varmuutta. Viimeisenä tavoitteena oli tekninen toteutus, joka suunniteltiin toteutettavaksi syksyllä 2011. (Savonia 2011.)

ISAT-tietovarastoprojektin projektisuunnitelman aikatauluun ei ollut aikataulutettu tietovarastoprojektissa toteutettavia tehtäviä, vaan ainoastaan keskitytty projektin päätavoitteisiin. Toisin sanoen aikatauluun oli merkattu päämäärät, joihin projektissa pyritään, mutta kantaa ei ollut otettu siihen, kuinka näihin päämääriin päästään. Aikatauluun olisi hyvä merkata päämäärien lisäksi toteuttavat tehtävät, eli se, mitä projektin aikana tulisi tehdä, että päämääriin päästään. Tehtävien aikataulutus on tärkeää, sillä kaikkia tehtäviä ei voida tehdä yhtä aikaa, vaan osa tehtävistä tulee tehdä vasta kun toinen tehtävä on valmistunut, eli kattoa ei voida asentaa ennen kuin seinät ovat pystyssä. Tehtävien erittely aikatauluun mahdollistaa myös projektin etenemisen seuraamisen, sillä tarkan aikataulutuksen ansiosta projektin tehtävistä tulee polku, jota seuraamalla projek-

ti saadaan toteutettua suunnitellusti, ilman epämiellyttäviä yllätyksiä. Tarkkaan aikatauluun on hyvä merkitä tehtävien lisäksi tehtävän tekijä, jotta tiedetään, mitä minkin henkilöresurssin on tehtävä ja missä vaiheessa kyseistä resurssia tarvitaan. (Hovi ym. 2009, 140.)

4.3.8 Budjetti

Viimeisenä lukuna projektisuunnitelmassa oli budjetointi. Budjetoinnin tarkoituksena on olla projektin taloudellinen suunnitelma joka on sidottu aikaan. Budjetin ei kuitenkaan tule olla sidoksissa kalenterissa esitettyyn aikaan, kuten kuukauteen tai vuoteen vaan projektin aikatauluun. Mikäli aikataulussa tapahtuu muutoksia, tulee projektin budjettiakin muuttaa niin, että se kattaa projektissa tapahtuvien muutokset taloudellisesti. Projektibudjetti on usein ns. kustannusbudjetti, eli budjetin tarkoituksena on kuvata ainoastaan projektista aiheutuvia kuluja ja niiden rahoittamista. Ennen kuin budjettia voidaan määrittää, tulee projektin tehtävien suoritusjärjestys, eli projektin vaiheistus, ja projekti aikataulu olla valmiit. (Pelin 2008, 175.)

ISAT -tietovaraston tietovarastoprojektissa budjetti jaettiin kolmeen osa-alueeseen. Nämä osa-alueet ovat liittymien toteutus, lisenssit ja palkat. Liittymien toteutuksella tarkoitetaan sitä rahallista summaa, joka on varattu operatiivisten tietojärjestelmien ja tietovaraston liittymien toteutukselle. Lisensseihin on budjetoitu ohjelmistojen, eli tietokantojen sekä BI-välineiden kustannukset. Palkkoihin kuuluu suunnittelun, toteutuksen ja ylläpidon kustannukset. (Savonia 2011.)

4.4 Vertailu

Tässä luvussa vertaillaan ISAT -tietovarastoprojektia yleisesti käytössä oleviin tietovarastojen toteutusmenetelmiin, jotka on esitetty luvuissa 3.5–3.9.

4.4.1 Tietolähteet

ISAT -tietovarastoon tuodaan tietoa pääosin PKAMK:n ja Savonian omista operatiivisista järjestelmistä, eikä ulkoisia lähteitä ole tarvittu tietovarastoa varten. Tämä johtuu siitä, että ISAT-tietovarasto on suunniteltu sisältämään ainoastaan opiskelijahallinnon, henkilöstöhallinnon ja taloushallinnon tietoa. Raportteja siis tehdään ainoastaan ammattikorkeakoulujen omista järjestelmistä tietovaraston ladatuista tiedoista. Ulkoiselle tiedolle ei raporteissa ole tarvetta alkuperäisten suunnitelmien mukaisesti. Tästä huolimatta ulkoisten tietolähteiden täysvaltaista sulkemista ei kannata koskaan tehdä, sillä josain tietovaraston kehitysvaiheessa saatetaan huomata, että ulkoiselle tiedolle on käyttöä. Tilanteita ulkoiselle tiedolle voivat olla esimerkiksi taloushallinnon tarve ulkoisille taloustiedoille, kuten erilaisille budjettitiedoille. (PKAMK 2011.)

Tietovarastoon siirtyvä tieto on strukturoitua tietoa, sillä tieto siirretään ammattikorkeakoulujen operatiivisista järjestelmistä tietovarastoon. Näihin operatiivisiin järjestelmiin tallennetaan ainoastaan strukturoitua tietoa, eli tietoa joka on mitattavissa. Tarkoituksena ei ole ottaa mukaan tietovarastoon strukturoimatonta tietoa, kuten sähköposteja tai dokumentteja. (PKAMK 2011.) Tätäkään vaihtoehtoa ei kannata sulkea pois, sillä paljon organisaatioille tärkeätä tietoa liikkuu muussa kuin strukturoidussa muodossa erilaisten tiedostojen kautta. Nämä tiedostot voivat sisältää sellaista tietoa, mitä ei voida siirtää operatiivisen järjestelmän kautta strukturoituun muotoon. Tällaisissa tapauksissa onkin erittäin hyödyllistä sallia strukturoimattoman tiedon vienti tietovarastoon, jolloin tiedostojen tietoa voidaan hyödyntää. (Hovi ym. 2009, 18.)

ISAT -tietovarastossa käytetyt tietolähteet ovat hyvin samantyyppisiä kuin Hovi yms. (2009, 18) kuvasivat, eli pääosin käytetään ammattikorkeakoulujen omia operatiivisia järjestelmiä tietolähteinä. Ainoa ero on ulkoisten tietolähteiden puute, joka selittyy sillä, että ammattikorkeakoulut eivät tunne tarvitsevansa raportteihinsa ulkoisia tietoja käytettäväksi. Tärkeintä on saada omat tiedot sellaiseen muotoon, että niistä on helppo tehdä raportteja ja tieto on sellaista, joka voidaan siirtää vaivattomasti Raketti-XDW-kantaan. Näihin tarkoituksiin valitut tietolähteet ovat erittäin hyviä.

4.4.2 Tietovarastotekniikka

ISAT-tietovarastossa on käytössä useita erilaisia tietovarastotyyppisiä, joita yhdistelmällä koko ISAT-tietovarastoarkkitehtuuri toteutetaan. Arkkitehtuuri koostuu keskistetyistä tietovarastosta, ammattikoulukohtaisista työalueista, käyttäjäryhmäkohtaisista datamarteista. Kaikki nämä tietovarastoarkkitehtuurin osat ovat samassa fyysisessä tietokannassa. Tietokantapalvelimena käytetään ISAT-tietovarastossa Microsoftin valmistamaa Microsoft SQL Server 2008 R2 -palvelinohjelmistoa. (PKAMK 2011.)

Arkkitehtuurin ideana on luoda vaiheittainen lataus, jonka jokaisessa vaiheessa tietoa muokataan sillä hetkellä tarvittavalla tavalla. Näin tietoa ei tarvitse muokata kokonaisuudessa yhden vaiheen aikana. Ensimmäisessä vaiheessa tiedot poimitaan PKAMK:n ja Savonian operatiivista järjestelmistä kummankin ammattikorkeakoulun omalle työalueelle. Tässä vaiheessa työalueelle siirretty tieto on vielä samassa muodossa, eli se vastaa pääsääntöisesti sen operatiivisen järjestelmän rakennetta, mistä kyseinen tieto on poimittu. Tietoa ei ole vielä muokattu millään tavalla, mutta operatiivisissa järjestelmissä on pyritty käyttämään yhteistä koodistoa niin paljon kuin se on mahdollista. Tämä tarkoittaa sitä, että operatiivisissa järjestelmissä on pyritty käyttämään samoja koodeja osoittamaan esimerkiksi sukupuolta tai kaupunkia. (PKAMK 2011.)

Latauksen toisessa vaiheessa tieto yhdenmukaistetaan työalueella, jolloin siitä tulee vertailukelpoista. Yhdenmukaistamisen jälkeen tieto siirretään keskitettyyn tietovarastoon, johon on tarkoituksena tuoda kaikki raportointiin tarvittava tieto säilytettäväksi. Siirrostai aiheutuneet virhetilanteet tallennetaan lokille, jotta tiedosta vastaava henkilö voi oikeista tiedon oikeaksi, että se ei aiheuta enää virhetilannetta latauksessa. Virhetilanteita aiheuttavaa tietoa ei siirretä tietovarastoon, vaan se jätetään operatiivisen järjestelmään odottamaan korjausta ja siirretään vasta seuraavassa ajossa, mikäli virhe on korjattu. Näin virheellinen tieto saadaan pidettyä pois keskitetystä tietovarastosta ja ennen kaikkea raporteista. Tämä virheiden tarkistus järjestelmä estää väärin tietojen pohjalta syntyviä päätöksiä. (PKAMK 2011.)

Latauksen viimeisessä vaiheessa tieto muokataan käyttäjän haluamaan muotoon BI-välineillä. Nämä tiedot saadaan paikallistietovarastoista, eli datamarteista. Datamarttien on tarkoitus tukea eri käyttäjäryhmien, kuten esimerkiksi opiskelijahallinnon ja henkilöstöhallinnon, raportointitarpeita. Raportteihin voidaan tuoda tietoja myös summata-

luista, joita muodostetaan keskitettyyn tietovarastoon ladatuista tiedoista. Summataulun ansiosta sein summausta tarvittavat tiedot ovat nopeasti käytössä, eikä summausta tarvitse tehdä uudelleen jokaista raporttia muodostaessa. Käyttäjakohtaisten datamarttien lisäksi keskitetystä tietovarastosta ladataan tietoa Raketti-XDW-kantaan, oppilaitosten ulkoista käyttöä varten. (PKAMK 2011.)

ISAT-tietovarastossa käytettävää arkkitehtuuria kutsutaan keskitetyksi yritystason tietovarastoksi eli EDW-arkkitehtuuriksi. Tämän arkkitehtuurin ideana on koota kaikki organisaation tiedot yhteen tai muutamaaan isoon tietovarastoon, jolloin tietovarasto pitää sisällään usean liiketoiminta-alueen tiedot yhdenmukaistettuna. ISAT-tietovarasto noudattaa erityisen tarkasti kyseisen arkkitehtuurin ideologiaa. Jopa niin pitkälle, että itse keskitetystä tietovarastosta ei tehdä lainkaan kyselyjä, vaan kaikki käyttäjien tekemät kyselyt ja raportointi tulee heille suunnitelluista datamarteista. Näistä Datamarteista pyritään tekemään mahdollisimman helppokäyttöiset sitä käyttäville henkilöille, jolloin vaikeammat ja teknisemmätkin kyselyt ja raportit saadaan tehtyä BI-välineillä ilman teknisten tietokantatyökalujen opiskelua. (Hovi ym. 2009, 27–28)

EDW -arkkitehtuurissa voidaan käyttää yhtä ainoaa palvelinta ja tietokantaa. Tietokantaratkaisu voidaan keskittää yhteen paikkaan, eikä sitä tarvitse levittää organisaation sisällä. Tämä tarkoittaa sitä, että luotuja datamartteja ei tarvitse asentaa esimerkiksi taloushallinnon ja opiskelijahallinnon omille palvelimille. Tätä ominaisuutta hyödynnetään ISAT-tietovaraston toteutuksessa, sillä suunnitelman mukaisesti arkkitehtuurin eri osia ei ole jaoteltu eri tietokantoihin. (PKAMK 2011.)

4.4.3 Tietovaraston suunnittelu

Tässä luvussa verrataan sitä, kuinka hyvin ISAT -tietovaraston suunnittelussa otettiin huomioon luvussa 3.7 mainitut tietovaraston suunnittelun kolme pääkohtaa. Nämä pääkohdat olivat tietovaraston laajuus, käyttäjien tarpeet ja saatavilla olevat tiedot.

Tietovaraston laajuutta käsiteltiin jo hieman edellisessä luvussa, sillä valittu arkkitehtuuri määrittää hyvin paljon sitä, kuinka laaja toteutettavasta tietovarastosta tulee. ISAT-hankkeen tietovaraston laajuutta määrittäessä ei ole tarvetta määrittää yhtä, tai useaa pientä kohdealuetta, sillä tietovaraston on tarkoitus palvella kahden oppilaitoksen kaik-

kea toimintaa. Tämä tarkoittaa sitä, että tietovaraston kohdealueena on koko organisaation toiminta. Tämän takia paras ratkaisu ISAT-tietovarastolle on tehdä EDW-arkkitehtuurin mukainen tietovarastokokonaisuus. (PKAMK 2011.)

EDW:n ansiosta ISAT-hankkeeseen osallistuvat oppilaitokset saavat pitkälle tulevaisuuteen palvelevan tietovarastokokonaisuuden, joka luo oppilaitoksille ns. datawarehouse-infrastruktuurin. Infrastruktuuri sisältää yhtenäisen tietovarastoympäristö, palvelimet, BI-, ETL- ja raporttivälineet yms. (Hovi ym. 2009, 28–29.) Tutkimusten mukaan tietovarastoprojektit, joissa tietovarastona käytetään EDW:tä onnistuvat suuremmalla todennäköisyydellä kuin muita arkkitehtuureja käyttäessä. Tämä osoittaa, että tietovarastot kannattaa suunnitella ja työstää tällä ratkaisulla, vaikka se aluksi tuntuu muita arkkitehtuuriratkaisuja työläämmältä. (Hovi ym. 2001, 70.)

ISAT-projektin tietovaraston laajuus vaikuttaa myös käyttäjien tarpeisiin. EDW:n ajatusmallin ansiosta käyttäjät saavat lähes kaiken järjestelmiin syötetyn tiedon käytettäväkseen raporteissaan. Tällöin käyttäjien tarpeita ei tarvitse huomioida tietovaraston laajuutta selvittäessä, sillä he saavat joka tapauksessa tarvitsemansa tietonsa keskitetystä tietovarastosta. Käyttäjien tarpeita palvellaan myös tarjoamalla mahdollisuus tehdä kyselyitä BI-välineillä raportointialuekohtaisista datamarteista. Ilman BI-välineitä tietovarasto olisi raportoinnin kannalta aivan samanlainen kuin operatiivinen järjestelmä, joten BI-välineiden ansiosta tietovarastosta saaduista raporteista tulee monipuolisempi kuin operatiivisten järjestelmien raporteista. (Hovi ym. 2009, 27–32) Käyttäjien tarpeet huomioidaan myös jättämällä virheellinen data pois tietovarastosta. Ilman tietovaraston virheentarkistusta, virheellistä dataa on mahdollista päästä operatiivisen järjestelmien tietokantoihin joka lopulta siirtyy raporteihin. Tietovaraston ansiosta käyttäjät voivat olla varmoja tiedon oikeellisuudesta. (PKAMK 2011.)

Tietovaraston laajuus vaikuttaa myös poimittavien tietojen valintaan. EDW-ratkaisussa yleinen ratkaisu onkin valita poimittavaksi kaikki operatiiviseen järjestelmiin syötetty tieto, jota jatkojalostetaan myöhemmin raportointikohtaisissa datamarteissa ja käyttäjien toimesta BI-välineillä. Näin voidaan olla varmoja, että kaikki käyttäjien tietotarpeet tyydytetään. Tämä malli on nykyisissä tietovarastoratkaisuissa erittäin yleinen, sillä levytila, jolle tietovarastoon poimittu tieto tallennetaan, on nykyään erittäin halpaa. Tietovarastoon ladattu tieto on karkeudeltaan erittäin hienojakoista, sillä kaikki tiedot siir-

retään tietovarastoon sellaisena, kuin se on syötettynä operatiiviseen järjestelmään. Tämä vaihtoehto on halvan levytilan takia erittäin yleinen ja suureen keskitettyyn EDW-varastoon tämä vaihtoehto on erittäin suosittu, sillä kyseiseen arkkitehtuuriin on tarkoitus tuoda kaikki yritykselle tarpeellinen tieto käyttäjien saataville. (Hovi ym. 2009, 27–34.) ISAT-tietovarastossa käytetään juuri tätä ideologiaa, eli kaikki opiskelijahallinnon, henkilöstö- ja taloushallinnon tiedot poimitaan keskitettyyn tietovarastoon (PKAMK 2011).

ISAT-hankkeen tietovarasto mallinnettiin juuri Hovin yms. (2009, 33–34) mainitsemanla tavalla. Ensin toteutettiin tietovaraston kokonaismalli, eli kokonaisvaltainen EDW-ratkaisu (PKAMK 2011). Tämän jälkeen suunniteltiin toteutettavaksi tietovaraston ensimmäinen osa-alue, eli opintohallinnon osuus tietovarastokantaan. Näin tietovarastosta saadaan tietty osa-alue käyttäjien käytettäväksi ennen tietovaraston lopullista valmistumista. Ensimmäisen osa-alueen valmistuttua aletaan suunnitella ja toteuttaa muita osa-alueita, eli tässä tapauksessa henkilöstö- ja taloushallinnon tietojen siirtoa. (Savonia 2011). Tietovaraston käyttöönotto pienissä erissä, eli inkrementeissä, on erittäin turvallinen valinta onnistumisen kannalta, sillä jokaisesta inkrementistä saadaan kokemusta ja palautetta käyttäjiltä, jolloin seuraavien osien toteuttaminen on nopeampaa. Inkrementaalinen toteutus lisää myös projektin uskottavuutta, sillä projektin johdolle saadaan näytettäväksi konkreettisia tuloksia jokaisen inkrementin jälkeen.

4.4.4 Tietojen summaaminen ja jalostaminen

Tietojen summaaminen ja jalostaminen nopeasti saatavaan muotoon on erittäin tärkeätä tietovarastoinnissa ja näitä tietovarastoinnille olennaisia osia ei ole jätetty pois ISAT-projektin tietovaraston toteutuksessa. ISAT-hankkeen tietovarastossa on varattuna summataulu, johon summauksia tehdään valmiiksi. Tämän taulun tarkoitus on tarjota tieto valmiiksi summattuna usein käytettyä tietoa ja tunnuslukuja. (PKAMK 2011.) Summataulu lisää myös tietovaraston suorituskykyä, sillä jokainen käyttäjän tekemä kysely vaatii tietovarastosta tietyn määrän suorituskykyä (Hovi ym. 2009, 45–47).

4.4.5 Tietojen reitit tietovarastoon

ISAT-hankkeen tietovarastoon tietoa poimitaan, muokataan ja ladataan ETL-arkkitehtuurin vetomenetelmää hyväksi käyttäen. Tiedot ladataan operatiivisista järjestelmistä tietovarastoon. Vetomenetelmässä ei käytetä siirtotiedostoja, joten minkäänlaisia rajapintaa ei luoda tietovaraston ja operatiivisen järjestelmän väliin kuten työntömenetelmässä, vaan tiedot poimitaan, muokataan ja ladataan operatiivisten järjestelmien tietokannoista suoraan. (PKAMK 2011.)

Vetomenetelmä toteutetaan käytetyn tietokantaratkaisun, eli Microsoft SQL Server 2008 R2:n, tietokantavälineillä (PKAMK 2011). Microsoftin SQL Serverissä käytettävä tietokantaväline on nimeltään SSIS eli SQL Server Integration Services. SSIS-välineellä voidaan poimia tietoa esimerkiksi tietokannoista, kuten SQL- ja Oracle-tietokannoista. Tietokantojen lisäksi SSIS:llä voidaan poimia tietoa tallennetuista tiedostoista, kuten Excel -taulukoista. Poiminnan jälkeen tietoa voidaan muokata halutulla tavalla ja lopuksi siirtää haluttuun kohteeseen tai kohteisiin. SSIS sisältää kaikki tarpeelliset toiminnot tiedon siirtämiseksi tietovarastoon. (McCown 2010.)

4.4.6 Yhteenveto

ISAT-projektin tietovaraston projektisuunnitelmassa otettiin huomioon samoja osa-alueita, joita tietovarastoinnin ammattilaiset suosittelevat ottamaan huomioon, kun tietovarastoa rakennetaan. Tästä huolimatta, projektisuunnitelmassa olisi voitu keskittyä eri osa-alueisiin tarkemmin, sillä nykyisessä projektisuunnitelmassa tehtiin vain hyvin pintapuolinen tarkastelu. Tarkka ja syvä luotavaa projektisuunnitelma estää riskien, kuten epäselvien ja epämääräisten tavoitteiden ja projektin tavoitteiden jatkuvan muuttumisen, toteutumista. (Hovi ym. 2009, 160).

Projektisuunnitelman pintapuolisesta tarkastelusta esimerkkinä voidaan pitää roolien jakamista. Projektisuunnitelmassa oli mainittu ainoastaan projektipäällikkö ja tilastojen, mittareiden ja tiedon oikeellisuudesta vastaavan vastuuhenkilön roolit. On tietenkin ymmärrettävää, että jokaista roolia ei voida aluksi nimetä, sillä osaamista hankitaan useissa tapauksissa ulkopuolisilta toimijoilta, mutta projektin alkuvaiheessa on erittäin

tärkeää nimetä projektissa työskentelevät henkilöt, jotta heillä on mahdollista valmistautua tehtäviinsä ja suoriutua niistä. (Hovi ym. 2009, 161.)

Tietovaraston toteutussuunnitelmassa otettiin hyvin huomioon keskitetyn tietovaraston ideologia, sillä suunnitelmassa esitetty tietovarasto oli esitetty täysin EDW-mallin mukaisesti. Latauksessa käytetty vetomenetelmässä on omat heikkoutensa, kuten se, että operatiivisen järjestelmää vaihtaessa joudutaan tekemään enemmän työtä latauksen liittämisessä uuteen järjestelmään, kuin työntömenetelmällä. Vetomenetelmän hyötyinä ovat yksinkertaisuus ja suoraviivaisuus, sillä tieto luetaan suoraan operatiivisen järjestelmä tietokannasta tietovarastoon. Vetomenetelmä on myös nykyaikaisempi ratkaisu. (Hovi ym. 2001, 85–86.)

5 Tietovaraston hyödyt ISATille

Tietovarastoa käytetään yrityksen operatiivisten toiminnan tukemiseen. Tietovarastoa käyttäessä operatiivisten järjestelmien rooli oppilaitoksissa muuttuu. Tietovarastoon keskitetään tieto yhteen paikkaan käyttäjien saataville raportointia ja analysointia varten. Operatiiviseen järjestelmään käytetään ainoastaan tiedon syöttämistä varten. Tämä vähentää huomattavasti operatiivisiin järjestelmiin kohdistuvaa kuormaa, sillä kyselyjä ei tarvitse tehdä operatiivisesta järjestelmästä. Operatiivisia järjestelmiä ei ole suunniteltu raskaaseen tietojen analysointiin ja raportointiin. Hyötynä kuorman vähenemisestä on operatiivisten järjestelmien toiminnan tehostuminen, joka näkyy operatiivisten järjestelmien käyttäjille järjestelmän nopeampana toimintana. Tietovarasto nopeuttaa myös kyselyjen ja raporttien tekemistä, sillä tietovarasto on ainoastaan tiedon lukemista varten ja se on suunniteltu erittäin raskaitakin kyselyitä varten. Ainoa kuorma mikä tietovarastoon kohdistuu, on tiedon käsittelevien käyttäjien tekemät kyselyt. Tietovarasto tukee niin, tietoja syöttävien, kuin tietoja käsittelevien käyttäjien toimintaa. (Törmänen 1999, 36.)

Tietovarastoa voidaan hyödyntää operatiivisten järjestelmien välisessä kommunikoinnissa. Tietovarastoon voidaan yhdistellä tietoja eri operatiivisten järjestelmien tietokannoista ja luoda rajapintoja ulkoisiin tietolähteisiin. (Törmänen 1999, 36.) Rajapinnan luonti ulkoiseen lähteeseen on erityisen tärkeää ISAT:ssa, sillä tietojen lataaminen ul-

koisiin tietolähteisiin oli yksi tavoitteista ISAT:n tietovarastossa. Ulkoisena tietolähteenä toimii Raketti-XDW, johon ladataan korkeakoulujen tietoja niin mukana olevien korkeakoulujen kuin opetusministeriön tarpeisiin. Liiketoiminnallisia etuna yhteiskäyttöön luodusta Raketti-XDW:stä on edullisemmat kustannukset. Palvelinalustan, ylläpidon ja tukipalveluiden kustannukset jakamalla korkeakoulut saavat tietovaraston käyttöön edullisemmin. Kustannusten jakautumisen lisäksi opetusministeriö tukee tietovaraston rakennus- ja kehitystyötä rahallisesti. (CSC 2012.)

ISAT:ssa mukana olevien oppilaitosten operatiivisten järjestelmien tietojen yhdistely ja jäsentely tietovaraston avulla voidaan tehdä yhteenveto, joka sisältää tietoa oppilaitosten eri toiminta-alueista. Tämä tarkoittaa sitä, että opiskelija-, henkilöstö-, ja taloushallinnon tietoja voidaan yhdistellä tarvittaessa, jolloin oppilaitosten toiminnasta saadaan laajempi näkökulma. Suuremman kokonaiskuvan oppilaitoksesta tarjoava tietovarastointi ehkäisee siiloutumisesta johtuvia ongelmia. (Törmänen 1999, 36–37.)

Erilaisten ja eri tarpeisiin tehtävien raporttien tekeminen helpottuu tietovaraston ansiosta. Tietovarastosta voidaan tehdä samanlaisia vakioraportteja, joita operatiivista järjestelmistä tehtäisiin. Nämä perusraportit ovat käytetyin raporttiryhmä ja raportteja saate-taankin tehdä päivittäin, joskus jopa useammin. Näillä raporteilla on useimmiten eniten käyttäjiä, sillä niitä on helppo käyttää ja ne tarjoavat tarpeelliset vastaukset suurimmalla osalla käyttäjistä. Räätelöidyt raportit ovat tarpeen mukaan tehtyjä täsmäraportteja. Täsmäraportit ovat hieman vaativampia kuin päivittäin ajettavat perusraportit, mutta kunhan täsmäraportti on kerran saatu valmiiksi ja ajettua, se on käytössä uudelleen tarpeen mukaan. Räätelöidyillä täsmäraporteilla on vähemmän käyttäjiä kuin vakioraportteilla.

Viimeisenä raporttiryhmänä on ns. ad hoc -raportit ja -kyselyt, eli yksittäisiin tapaksiin tarkoitetut raportit ja kyselyt. Ad hoc -raportit tarvitaan ongelmatilanteiden selvittämiseen tai yksittäisen tuotantotapauksen selvitykseen. Tällä raporttityypillä on vähiten käyttäjiä, sillä ad hoc -raportit edellyttävät analyysi- ja metadatan hallintaa. (Törmänen 1999, 44.) Tietovaraston ansiosta oppilaitokset saavat tietovaraston ansiosta luotuja raportteja, joita operatiivisilla järjestelmillä ei voitaisi tehdä, kuten esimerkiksi ad hoc -raportit. Tietovaraston tarjoamien raporttien ansiosta päällekkäiset työt vähenevät, sillä tietoja ei tarvitse enää etsiä ja kerätä eri lähteistä. Tietovarastoa voidaankin kutsua ra-

porttivarastoksi, josta kaikki oppilaitoksen tarvitsemat raportit ovat saatavilla helposti ja nopeasti. (Hovi ym. 2001, 150.)

Tietovarasto tukee innovaatioita ja mahdollistaa uusien liiketoiminta-alueiden löytämisen. Tässä tapauksessa innovaation ei tarvitse olla uusi teknologinen idea. Innovaatio on uudenlainen näkemys tehdä asioita, toisin sanoen, järkevämpi ja kustannustehokkaampi tapa toteuttaa yrityksen tehtäviä. Innovaatioita ja vielä tuntemattomia liiketoiminta-alueita voidaan löytää yritysten omasta tietomassasta ja tuon tietomassan tarkasta hallinnasta. Tietomassa hallinnan ansiosta oppilaitokset tuntevat omat organisaationsa tarkemmin, jolloin uusien innovaatioiden keksiminen ja hyödyntäminen on yksinkertaisempaa. Tietovaraston avulla oppilaitokset voivat tehdä tiedon louhintaa, eli etsiä jo olemasta olevasta tiedosta yhtäläisyyksiä ja poikkeavuuksia. Etsinnän jälkeen tietoa aletaan tulkita ja tarkastella eri näkökulmasta. (Törmänen 1999, 42–43.)

Tietovarastot maksavat itsensä takaisin hyvin nopeasti. IDC:n tekemän tutkimuksen mukaan tietovaraston toteuttamisesta aiheutuvat kulut on katettu keskimäärin 2,3 vuodessa. Tietovarastoon käytetty aika ja raha palautuvat hyvin nopeasti ja kulujen kattamisen jälkeen tietovarasto tuo oppilaitoksille ainoastaan voittoa. Tietenkään tämä ei tarkoita, että tietovaraston toteuttaminen ei ole riskitöntä, mutta hyvän suunnittelun avulla varmistetaan, että tietovarasto ei ainakaan aiheutua mittavia tappioita. (Kelley, 2001.)

6 Pohdinta

Valittu opinnäytetyön aihe oli mielenkiintoinen, sillä pidän tietovarastointia erittäin kiinnostavana aiheena ja olen aina halunnut oppia aiheesta enemmän. Kuultuani ISAT-projektista ja siinä toteutettavasta tietovarastosta opinnäytetyön ohjaajaltani ajattelin, että kyseinen aihe on erittäin sopiva opinnäytettä varten, sillä projektin aikana pääsi tutustumaan toteutettavan tietovaraston suunnitteluun ja toteuttamiseen.

Tiedon kerääminen opinnäytettä varten osoittautui erittäin helpoksi, sillä tietovarastoinnista on kirjoitettu paljon. Kirjallisuutta on suomeksi ja useammilta vuosilta, jolloin tietovarastoinnin kehitystä pääsi seuramaan. Hankalin osuus opinnäytetyössä oli itse kirjoittamisen aloittaminen, sillä aiheen rajaaminen ja tiedon kerääminen ISAT-

projektista osoittautui suunniteltua vaikeammaksi. Projektin eteneminen oli hidasta, joten tietoa projektin eri vaiheista ei ollut käytettävissä silloin kun sitä tarvitsi, mikä aiheutti sen, että opinnäytetyön kirjoittamisen aloittaminen ei mennyt suunnitellun aikataulun mukaisesti. Ongelmista huolimatta opinnäytetyön kirjoituksen lähdettyä kunnolla käyntiin ei ongelmia enää ilmennyt, sillä olin ehtinyt hahmottaa, mitä asioita opinnäytetyössä tulisi käsitellä.

Tein rajauksen vertailussa niin että toiminnallisessa osuudessa pyrin vertailemaan ISAT-projektin tietovaraston projekti- ja toteutussuunnitelmaa yleisesti käytettäviin menetelmiin ja toimintamalleihin. Valitsemani rajausta osoittautui hyväksi, sillä juuri näihin kahteen osa-alueeseen tuli eniten tietoa projektin aikana. Aiheen rajaaminen on erittäin tärkeä tietovarastoinnista opinnäytetyötä tehdessä, muutoin opinnäytteestä puuttui selkeä käsittelykohde ja opinnäytteestä tulee erittäin laaja. Tietovarastoinnissa voidaan mennä hyvinkin teknisiin asioihin, joten pyrin käsittelemään tietovarastoinnin perusteita mahdollisimman selkeästi.

Projektin materiaalin pohjalta aloin selvittää, kuinka materiaalissa esitetyt menetelmät ja toimintamallit poikkeavat kirjallisuuden vastaavista. Näin sain kokonaiskuvan meneillään olevasta projektista ja ymmärsin, miksi siinä on toimittu valituilla tavoilla. Projektista muodostuneen kokonaiskuvan lisäksi aloin hahmottamaan tarkemmin mitä tietovarastointi on, ja miksi sitä tarvitaan. Opinnäytetyössä oppimani asiat tietovarastoinnista kasvattavat ammatillista osaamista ja lisäävät ymmärrystäni tietokannoista ja tietovarastoista. Tämä oli yksi tärkeimmistä itselleni asetetuista tavoitteista, sillä olen ollut kiinnostunut tietovarastoista niistä kuultuani koulutuksessani. Ammattini kannalta opinnäyte on minulle erittäin hyödyllinen. Uskon myös, että se on hyvä katsaus ISAT-projektin tietovaraston toteutukseen projektin ulkoisesta näkökulmasta. Ulkoisesta näkökulmasta on varmasti hyötyä toimeksiantajalleni, sillä omalle työlle voi tulla sokeaksi ja uuden tuoreen näkökulman myötä omaan työhön alkaa suhtautua eri tavalla.

Lähteet

- Kelley, C. 2001. The IDC data warehousing ROI study: An analysis. TechTarget.
<http://searchsqlserver.techtarget.com/tip/The-IDC-data-warehousing-ROI-study-An-analysis>. 8.10.2012.
- CSC. 2012. Raketti-XDW. Tieteen tietotekniikka keskus.
<http://raketti.csc.fi/paattyneet/xdw>. 9.7.2012.
- Hovi, A. 1997. Data Warehousing - Tietovarastotekniikka. Espoo: Suomen Atk-kustannus Oy.
- Hovi, A. Hervonen, H & Koistinen, H. 2009. Tietovarastot ja Business Intelligence. Jyväskylä: WSOYpro/Docendo-tuotteet.
- Hovi, A. Koistinen, H. Ylinen, J. 2001. Tietovarastot liiketoiminnan tukena. Espoo: Satku – Kauppakaari.
- Kerr, T. 2003. ETL in a Box. Information Management. <http://www.information-management.com/issues/20031101/7607-1.html> 10.9.2012.
- McCown. 2010. Introduction to the SQL Server Integration Services (SSIS) Development Environment. <http://www.petri.co.il/sql-server-integration-services-introduction.htm>. 13.08.2012.
- Nishith. 2006. Data Mart vs Data Warehouse – The Great Debate. Open Source Analytics. <http://opensourceanalytics.com/2006/03/14/data-mart-vs-data-warehouse-the-great-debate/>. 29.02.2012.
- PKAMK. 2011. ISATDW -tietovaraston toteutussuunnitelma. Kirmanen, Mönkkönen. 15.3.2011.
- PKAMK. 2012a. Info. Pohjois-Karjalan ammattikorkeakoulu.
http://www.ncp.fi/index.php?option=com_content&view=article&id=49&Itemid=78. 26.1.2012.
- PKAMK. 2012b. ISAT – Itä-Suomen Ammattikoulut. Itä-Suomen Ammattikorkeakoulut. <http://www.isat.fi/>. 30.1.2012.
- Pelin, R. 2008. Projektihallinnan käsikirja. Jyväskylä: Projektijohtaminen Oy Risto Pelin.
- Savonia. 2011. ISAT-tietovarasto-projektisuunnitelma. Savonia-ammattikorkeakoulu. 14.3.2011.
- Savonia. 2012a. Tutustu Savoniaan. Savonia. <http://portal.savonia.fi/amk/tutustu-savoniaan>. 30.1.2012.
- Savonia. 2012b. Nuorten AMK-tutkinnot. Savonia.
<http://portal.savonia.fi/amk/hakijalle/amk-ja-yamk-tutkinnot/nuorten-amk-tutkinnot>. 30.1.2012.
- Savonia. 2012c. ISAT-Ristiinopiskelu. Savonia.
<http://portal.savonia.fi/amk/hakijalle/isat-ristiinopiskelu>. 30.1.2012.
- Sininen Meteoriiitti. Esitutkimus ja Strateginen suunnittelu. Sininen Meteoriiitti.
<http://www.meteoriiitti.com/fi-FI/palvelut/konsultointipalvelut/esitutkimus-ja-strateginen-suunnittelu/>. 3.6.2012.
- Törmänen, A. 1999. Tietovarastointi – strategiasta toteutukseen. Helsinki: Suomen Atk-kustannus Oy.
- Valtiovarainministeriö. 2011. JulkICT-toiminto – tietohallintolaki. Valtiovarainministeriö.
http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/Tieto_hallintolaki-esite.pdf. 25.1.2012.
- Wikipedia. 2011. Savonia-ammattikorkeakoulu. Wikipedia.
<http://fi.wikipedia.org/wiki/Savonia-ammattikorkeakoulu>. 30.1.2012.